PROFICIENCY OF NERVE CONDUCTION USING STANDARD METHODS AND REFERENCE VALUES (Cl. NPhys TRIAL 4)

WILLIAM J. LITCHY, MD,[1] JAMES W. ALBERS, MD,[2] JAMES WOLFE,[2] CHARLES F. BOLTON, MD,[3] NANCY WALSH, R ET, RT (EMG),[3] CHRISTOPHER J. KLEIN, MD,[1] ANDREW J. ZAFFT, R EMG/EP T,[1] JAMES W. RUSSELL, MD,[4] MELANIE ZWIRLEIN, R EMG/EEG/EPT,[4] CAROL J. OVERLAND,[1] JENNY L. DAVIES, BA,[1] RICKEY E. CARTER, PhD,[5] PETER J. DYCK, MD,[1] and the Cl. Nphys Trial 4 Investigators

[1] Department of Neurology, Mayo Clinic, 200 First Street SW, Rochester, Minnesota 55905, USA
[2] Department of Neurology, University of Michigan, Ann Arbor, Michigan, USA
[3] Department of Neurology, Queen's University, Kingston, Ontario, Canada
[4] Department of Neurology, University of Maryland, Baltimore, Maryland, USA
[5] Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, Minnesota, USA

Accepted 16 March 2014

ABSTRACT: *Introduction:* The Cl. NPhys Trial 3 showed that attributes of nerve conduction (NC) were without significant intraobserver differences, although there were significant interobserver differences. *Methods:* Trial 4 tested whether use of written instructions and pretrial agreement on techniques and use of standard reference values, diagnostic percentile values, or broader categorization of abnormality could reduce significant interobserver disagreement and improve agreement among clinical neurophysiologists. *Results:* The Trial 4 modifications markedly decreased, but did not eliminate, significant interobserver differences of measured attributes of NC. Use of standard reference values and defined percentile values of abnormality decreased interobserver disagreement and improved agreement of judgment of abnormality among evaluators. Therefore, the same clinical neurophysiologist should perform repeat NCs of therapeutic trial patients. *Conclusions:* Differences in interobserver judgment of abnormality decrease with use of common standard reference values and a defined percentile level of abnormality, providing a rationale for their use in therapeutic trials and medical practice.

Muscle Nerve **50**: 900–908, 2014

The need for the Cl. NPhys Trial 4 arose from consideration of the results of Cl. NPhys 3.[1] In Trial 3, 4 expert clinical neurophysiologists and their associate technologists evaluated 8 attributes of nerve conduction (NC) of the leg on 2 consecutive days in 24 masked patients without and with diabetic polyneuropathy. Although their intraobserver agreement was high and not significantly different between days 1 and 2, statistically significant interobserver differences, sometimes of some magnitude, were observed for most attributes of NCs tested. A similar result had been reported earlier by clinical neurophysiologists at Johns Hopkins.[2] Trial 4, similar in design to Cl. NPhys Trial 3, tests whether interobserver differences can be improved or eliminated by standardization of nerve conduction technique, use of common reference values, and setting abnormality at a given percentile value.

METHODS

Conduct of Cl. NPhys Trial 4. The design of Cl. NPhys Trial 4 was similar to that of Trial 3 (same clinical neurophysiologists, same or similar diabetic patients without and with diabetic polyneuropathy, assessment of the same 8 attributes of nerve conduction of the leg, masked evaluation of subjects, and preconditions of testing). In Trial 3, no directions were provided on how to test or judge abnormality of the 8 attributes of NC. In Trial 4, specific testing routines and methods of assessment were provided in a specially prepared syllabus and in a training session.

The same clinical neurophysiologists who had participated in Trial 3 performed the NC tests in Trial 4. They came from 4 different North American medical centers (London, Ontario, Canada; Baltimore, Maryland, USA; Ann Arbor, Michigan, USA; and Rochester, Minnesota, USA). Rather than have the 24 subjects travel from Rochester to the other medical centers for their NC evaluations on 2 consecutive days, we had the smaller group of investigators travel to Rochester, which allowed for a more rigorous and masked trial and a cost savings. Trial 4 patients without and with diabetic polyneuropathy included patients from Trial 3;

Abbreviations: CMAP, compound muscle action potential; EMG, electromyogram; MNCV, motor nerve conduction velocity; MNDL, motor nerve distal latency; NC, nerve conduction; NIS, Neuropathy Impairment Score; RDNS-HS, Rochester Diabetic Neuropathy Study—Healthy Subjects; SNAP, sensory nerve action potential; SNCV, sensory nerve conduction velocity

Key words: multicenter trial; nerve conduction accuracy; nerve conduction proficiency; neuropathy impairment score; reference values

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors have full access to all the data and the right to publish any and all data, separate and apart from the guidance of any sponsor.

Disclosure: P.J.D. serves as an Associate Editor for *Diabetes* and receives an honorarium. His laboratory receives financial support from the Mayo Foundation; Pfizer, Inc.; Isis, Inc.; and Alnylam, Inc.; and previously from the National Institutes of Health (NS 36797), FDA, and other pharmaceutical companies.

This study was supported in part by Mayo Foundation Funds and grants obtained from the National Institute of Neurological Disorders and Stroke (R01-NS36797 to P.J.D.) and the National Institute on Aging (R01-AG34676 to W.A.R.).

Correspondence to: P.J. Dyck; e-mail: dyck.peter@mayo.edu

Published online 19 March 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/mus.24243

others were recruited from the Rochester Diabetic Polyneuropathy Study, substituting for patients who had died or who no longer agreed to participate. Two of the 24 patients failed "at the last minute" to participate in Trial 4: 1 because he was snowbound and the other due to intercurrent illness.

Before Trial 4, examiners were sent a syllabus (prepared by W.J.L.) that detailed the specific procedures for the NC studies. The syllabus included descriptions of the requirements for the electromyographic (EMG) instrument; the equipment to be used, including materials for skin preparation, stimulating electrodes, and recording electrodes; and the methods to standardize recording and measurement of the attributes of NC to be evaluated. The NC attributes analyzed were: compound muscle action potential (CMAP) recorded with stimulation at the knee; nerve conduction velocity (MNCV) calculated from the CMAPs recorded with stimulation at the knee and ankle; motor distal latency (MNDL) from the CMAP with ankle stimulation; and sensory nerve action potential amplitude (SNAP) and sensory distal latency (SNDL) from the sural SNAP with stimulation 14.0 cm proximal to the active recording electrode. The 8 NC attributes evaluated were fibular and tibial MNCV, CMAP, and MNDL, and sural SNAP and SNDL.

The training syllabus and the clinical demonstration emphasized skin preparation; precise and standard electrode placement of stimulating, recording, and reference electrodes; techniques to provide optimal stimulation; use of just supramaximal stimulation; recognition of anomalous innervation; description of averaging for certain studies; and standard techniques to measure interelectrode distances and marking of the evoked responses (e.g., baseline to initial negative peak amplitude measurements).

On the day before the start of Cl. NPhys 4, investigators met to review the use of these standard NC techniques. One of us (W.J.L.) demonstrated the specific technique to be followed. All clinical neurophysiologists agreed to follow the exact procedures and techniques outlined in the syllabus. Standard EMG instruments [2 Nicolet EDX Viking (Middleton, Wisconsin) and 1 Cadwell Sierra Wave (Kennewick, Washington)] and peripheral materials were made available for the trial.

On the days of Trial 4 (May 2 and 3, 2013) as in Trial 3, the clinical neurophysiologists and their associate technologist occupied separate cubicles in a large hall. Each subject's legs were carefully pre-warmed in warm water, surface temperatures were measured between evaluations, and limbs were kept warm between evaluations using specially fabricated thermal casts. For NC testing, the lower limbs were introduced through a curtain into the examining cubicle so that the neurophysiologists could not identify the patient by sight. In addition, subjects were asked not to speak to examiners to maintain masking. Each neurophysiology team assessed 8 attributes of NC of the left leg of each of the 22 patients without and with diabetic polyneuropathy, made judgments of normality or abnormality of each attribute assessed, and made an overall judgment of whether the patient had electrophysiologic evidence of polyneuropathy. On the second day, the order of patient assessment was altered, but each of the 4 clinical neurophysiologists assessed the 8 attributes of NC of the 22 patients by performing the same testing procedures and making the same judgments as those on day 1. All visible marks of testing were removed between NC testing of different investigators.

Assessment of Intra- and Interobserver Differences and Agreement. Wilcoxon signed-rank tests were used to test for differences between days 1 and 2. Friedman χ^2 or Cochran Q tests were used to assess for interobserver differences. The Krippendorff α was used to assess interobserver agreement. Because there is considerable evidence that neurophysiological functions, such as hearing, vision, and somatic sensation, increase as linear exponential functions, and other neurophysiologic functions may increase similarly, we evaluated interobserver variability using raw neurophysiologic measurements in addition to broader and perhaps more clinically meaningful categories of normality and abnormality. To do this, measured NC data were transformed into normal deviates from percentile values as corrected for applicable physical variables, and they were also transformed into Neuropathy Impairment Score (NIS) points, namely attribute values $>$5th percentile = 0 point (normal), $\leq$5th to $>$1st = 1 point (abnormal), and $\leq$1st = 2 points (highly abnormal). Abnormality of latencies is expressed in the other tail of the normal distribution using the same "points from percentiles" categorization.

Comparison of Judgment of Polyneuropathy within Trial 4 and between Trials 3 and 4. Cl. NPhys 4 compared the frequency of the clinical neurophysiologists' judgment of polyneuropathy to transformations of their measured values to percentiles and set abnormality of 2 attributes commonly affected in diabetic polyneuropathy (fibular CMAP and sural SNAP) at $\leq$2.5th percentile.

The frequencies of interobserver differences of raw measurements of attributes of NC and frequencies of abnormality based on a percentile abnormality from standard reference values or NIS

Table 1. Raw values of measured attributes of nerve conduction: interobserver disagreement and agreement (and intraobserver disagreement*).

Nerve conduction attribute	Cl. NPhys	Day 1 Median	Day 1 Range	Day 2 Median	Day 2 Range	Friedman χ^2 test for differences among 4 Cl. NPhys teams: Day 1 χ^2	Day 1 P	Day 2 χ^2	Day 2 P	Krippendorff α (ratio): Day 1 α	Day 1 Bootstrap 95% CI	Day 2 α	Day 2 Bootstrap 95% CI
Fibular CMAP (mV)	1	2.0	0.0–8.0	1.6	0.0–8.3	32.02	<0.01	21.77	<0.01	0.75	(0.65–0.85)	0.81	(0.72–0.89)
	2	2.0	0.0–8.1	3.3	0.0–8.9								
	3	1.6	0.0–8.1	1.5	0.0–7.8								
	4	1.8	0.0–9.1	2.1	0.0–8.4								
Fibular MNCV (m/s)	1	41.0	35.0–48.0	43.0	35.0–49.0	3.67	0.29	11.00	0.02	0.81		0.81	
	2	40.5	30.0–49.0	41.0	33.0–49.0						(0.70–0.89)		(0.76–0.86)
	3	41.0	25.0–50.0	40.0	31.0–50.0								
	4	40.0	32.0–50.0	40.5	33.0–47.0								
Fibular MNDL (ms)	1	4.6	3.4–7.9	4.6	3.6–8.0	13.04	<0.01	4.87	0.18	0.87		0.84	
	2	4.8	3.7–8.0	4.6	3.8–6.6						(0.83–0.91)		(0.79–0.87)
	3	4.8	4.0–9.0	4.7	3.9–6.7								
	4	4.8	3.7–10.7	4.6	3.5–8.5								
Tibial CMAP (mV)	1	4.1	0.0–8.5	3.6	0.0–10.6	20.01	<0.01	20.41	<0.01	0.76		0.83	
	2	4.2	0.0–9.2	4.3	0.0–9.5						(0.65–0.86)		(0.75–0.90)
	3	3.7	0.0–9.4	3.9	0.0–8.0								
	4	4.2	0.0–11.9	4.3	0.0–9.1								
Tibial MNCV (m/s)	1	41.5	32.0–48.0	41.0	33.0–53.0	2.51	0.47	3.05	0.38	0.59		0.58	
	2	41.0	30.0–48.0	40.0	33.0–50.0						(0.46–0.71)		(0.44–0.70)
	3	41.0	34.0–48.0	41.0	31.0–47.0								
	4	40.0	33.0–52.0	41.0	36.0–49.0								
Tibial MNDL (ms)	1	4.4	3.5–6.8	4.1	3.7–6.3	31.43	<0.01	31.74	<0.01	0.39		0.45	
	2	4.5	3.6–5.8	4.2	3.6–5.8						(0.20–0.56)		(0.29–0.59)
	3	5.0	3.8–8.5	5.0	3.9–6.0								
	4	4.9	3.8–7.5	4.8	3.7–6.0								
Sural SNAP (μV)	1	3.5	0.0–9.0	3.5	0.0–10.0	23.83	<0.01	27.44	<0.01	0.62		0.65	
	2	5.0	0.0–11.0	4.5	0.0–11.0						(0.49–0.74)		(0.53–0.77)
	3	1.5	0.0–8.0	3.0	0.0–8.0								
	4	3.0	0.0–9.0	3.0	0.0–11.0								
Sural SNDL (ms)	1	4.1	3.3–5.0	4.0	3.5–5.3	3.86	0.28	0.68	0.88	0.85		0.87	
	2	3.9	3.4–4.8	4.0	3.3–4.7						(0.80–0.88)		(0.83–0.90)
	3	3.9	3.3–5.3	4.0	3.5–5.1								
	4	4.2	3.3–4.9	4.1	3.3–4.9								

**Wilcoxon signed-rank tests were not statistically significant between days 1 and 2.*

Shaded boxes represent statistically significant differences among the neurophysiologist results. Shaded boxes indicate p values ≤0.05.

point transformations (from percentiles) were compared between Trials 3 and 4.

RESULTS

Raw Values of Measured Attributes of NC: Interobserver Disagreement and Agreement and Intraobserver Disagreement. Table 1 shows the median and range of raw values of the 8 measured attributes of NC assessed in 22 subjects by the 4 groups of clinical neurophysiologists. Using the Friedman χ^2 test, statistically significant interobserver differences were found for 10 of 16 comparisons (shown as shaded boxes in Table 1). The degree of agreement among clinical neurophysiologists was assessed using the Krippendorff α (e.g., values ⩾0.75 were observed for 10 of 16 evaluations).

Using the Wilcoxon signed-rank test, no statistically significant intraobserver difference was found between days 1 and 2 (footnote to Table 1). These results show that, although interobserver differences were observed commonly, there were no significant intraobserver differences between days 1 and 2.

Attributes of NC Transformed to NIS Points from Percentiles (>5th/<95th = 0; ≤5th to >1st/≥95th to <99th = 1; and ≤1st/≥99th = 2) Percentiles Corrected for Applicable Variables from a Standard Reference Source [Rochester Diabetic Neuropathy Study—Healthy Subjects (RDNS-HS)]. Statistically significant interobserver disagreement was found for only 3 of 16 comparisons. Using this transformation, in Table 2 agreement among clinical neurophysiologists, as measured by the Krippendorff α, averaged 0.72 ± 0.19, very similar to the value obtained when

Table 2. Measured attributes of nerve conduction transformed to NIS points* from RDNS-HS† percentiles: interobserver disagreement and agreement (and intraobserver disagreement‡).

Nerve conduction attribute	Cl. NPhys	Day 1		Day 2		Friedman χ^2 test for differences among 4 Cl. NPhys teams				Krippendorff α (ratio)			
						Day 1		Day 2		Day 1		Day 2	
		Median	Range	Median	Range	χ^2	*P*	χ^2	*P*	α	Bootstrap 95% CI	α	Bootstrap 95% CI
Fibular	1	0.0	0.0–2.0	0.0	0.0–2.0	4.71	0.19	3.00	0.39	0.94	(0.89–0.99)	0.99	(0.98–1.00)
CMAP	2	0.0	0.0–1.0	0.0	0.0–1.0								
(points:	3	0.0	0.0–2.0	0.0	0.0–2.0								
3 categories)	4	0.0	0.0–1.0	0.0	0.0–1.0								
Fibular	1	0.0	0.0–2.0	0.0	0.0–2.0	7.60	0.06	3.24	0.36	0.77	(0.69–0.86)	0.79	(0.70–0.87)
MNCV	2	0.0	0.0–2.0	0.0	0.0–2.0								
(points:	3	1.0	0.0–2.0	0.0	0.0–2.0								
3 categories)	4	0.0	0.0–2.0	0.0	0.0–2.0								
Fibular	1	0.0	0.0–2.0	0.0	0.0–2.0	1.29	0.73	3.97	0.26	0.77	(0.63–0.90)	0.67	(0.50–0.82)
MNDL	2	0.0	0.0–2.0	0.0	0.0–2.0								
(points:	3	0.0	0.0–2.0	0.0	0.0–2.0								
3 categories)	4	0.0	0.0–2.0	0.0	0.0–2.0								
Tibial	1	0.0	0.0–2.0	0.0	0.0–1.0	4.00	0.26	6.33	0.10	0.99	(0.98–1.00)	0.94	(0.87–0.99)
CMAP	2	0.0	0.0–2.0	0.0	0.0–2.0								
(points:	3	0.0	0.0–2.0	0.0	0.0–2.0								
3 categories)	4	0.0	0.0–2.0	0.0	0.0–2.0								
Tibial	1	0.0	0.0–2.0	0.0	0.0–2.0	0.86	0.84	2.37	0.50	0.69	(0.59–0.78)	0.54	(0.38–0.69)
MNCV	2	0.0	0.0–2.0	0.0	0.0–2.0								
(points:	3	1.0	0.0–2.0	0.0	0.0–2.0								
3 categories)	4	0.0	0.0–2.0	0.0	0.0–2.0								
Tibial	1	0.0	0.0–2.0	0.0	0.0–2.0	16.10	<0.01	17.94	<0.01	0.31	(0.15–0.47)	0.48	(0.32–0.63)
MNDL	2	0.0	0.0–2.0	0.0	0.0–2.0								
(points:	3	1.0	0.0–2.0	1.0	0.0–2.0								
3 categories)	4	1.0	0.0–2.0	0.0	0.0–2.0								
Sural	1	0.5	0.0–1.0	0.5	0.0–1.0	7.76	0.05	8.59	0.04	0.70	(0.59–0.80)	0.69	(0.58–0.80)
SNAP	2	0.0	0.0–2.0	0.0	0.0–2.0								
(points:	3	1.0	0.0–2.0	0.5	0.0–2.0								
3 categories)	4	0.0	0.0–2.0	0.0	0.0–2.0								
Sural	1	0.0	0.0–2.0	0.0	0.0–2.0	4.13	0.25	3.02	0.39	0.59	(0.42–0.74)	0.73	(0.60–0.85)
SNDL	2	0.0	0.0–2.0	0.0	0.0–2.0								
(points:	3	0.0	0.0–2.0	0.0	0.0–2.0								
3 categories)	4	0.0	0.0–2.0	0.0	0.0–2.0								

**Percentile values corrected to NIS points: >5th = 0; ≤5th to >1st = 1; and ≤1st = 2 points, and similar conversions if values are in the upper tail of the distribution (e.g., motor nerve distal latencies).*

†Rochester Diabetic Neuropathy-Healthy Subjects (RDNS-HS) used to produce standard reference values.

‡Wilcoxon signed-rank tests were not statistically significant between days 1 and 2.

Shaded boxes represent statistically significant differences among neurophysiologist results. Shaded boxes indicate p values ≤0.05.

measured raw values were compared (previous subsection). Values ⩾0.75 were observed 7 of 16 times.

No significant intraobserver difference was observed between days 1 and 2 (footnote to Table 2).

Clinical Neurophysiologist Judgment of Abnormality of Attributes of NC: Interobserver Disagreement and Agreement and Intraobserver Disagreement. Statistically significant disagreement of attributes of NC was observed for 7 of 16 such comparisons (Table 3). The average Krippendorff α among clinical neurophysiologists for the 8 attributes of NC was 0.65 ± 0.15. High values (i.e., ⩾0.75) were observed for 5 of 16 comparisons. No intraobserver difference was found between days 1 and 2 (footnote to Table 3).

Judgment of Abnormality of Attributes of NC Using Defined Percentile Abnormalities (≤2.5th/≥97.5th) from a Standard Reference Source (RDNS-HS). Clinical neurophysiologist measurements of attributes of NC were converted to percentiles and declared to be normal or abnormal by the criterion of ≤2.5th/⩾97.5th based on healthy subject reference values from the RDNS-HS cohort. Using this criterion, statistically significant interobserver differences were observed in 4 of 16 such comparisons (Table 4). If the percentile criterion was changed to

Table 3. Clinical neurophysiologists' judgment of abnormality of attributes of nerve conduction: interobserver disagreement and agreement (and intraobserver disagreement*).

Nerve conduction attribute	Cl. NPhys	Number (%) of 24				Cochran Q test for differences among 4 Cl. NPhys teams				Krippendorff α (ordinal)			
		Day 1		Day 2		Day 1		Day 2		Day 1		Day 2	
		Normal	Abnormal	Normal	Abnormal	Q	*P*	Q	*P*	α	Bootstrap 95% CI	α	Bootstrap 95% CI
Fibular	1	11 (50.0)	11 (50.0)	11 (50.0)	11 (50.0)	4.80	0.19	4.71	0.19	0.85	(0.74–0.94)	0.90	(0.80–0.98)
CMAP	2	13 (59.1)	9 (40.9)	12 (54.5)	10 (45.5)								
abnormality	3	11 (50.0)	11 (50.0)	10 (45.5)	12 (54.5)								
(Dr.'s judgment)	4	13 (59.1)	9 (40.9)	12 (54.5)	10 (45.5)								
Fibular	1	13 (59.1)	9 (40.9)	14 (63.6)	8 (36.4)	13.55	<0.01	13.80	<0.01	0.55	(0.39–0.72)	0.62	(0.46–0.76)
MNCV	2	8 (36.4)	14 (63.6)	8 (36.4)	14 (63.6)								
abnormality	3	15 (68.2)	7 (31.8)	12 (54.5)	10 (45.5)								
(Dr.'s judgment)	4	15 (68.2)	7 (31.8)	15 (68.2)	7 (31.8)								
Fibular	1	17 (77.3)	5 (22.7)	18 (81.8)	4 (18.2)	1.94	0.58	2.40	0.49	0.59	(0.37–0.78)	0.75	(0.57–0.90)
MNDL	2	18 (81.8)	4 (18.2)	19 (86.4)	3 (13.6)								
abnormality	3	17 (77.3)	5 (22.7)	17 (77.3)	5 (22.7)								
(Dr.'s judgment)	4	19 (86.4)	3 (13.6)	18 (81.8)	4 (18.2)								
Tibial	1	13 (59.1)	8 (36.4)	14 (63.6)	8 (36.4)	6.00	0.11	6.00	0.11	0.83	(0.72–0.93)	0.90	(0.80–0.98)
CMAP	2	13 (59.1)	9 (40.9)	14 (63.6)	8 (36.4)								
abnormality	3	15 (68.2)	7 (31.8)	14 (63.6)	8 (36.4)								
(Dr.'s judgment)	4	16 (72.7)	6 (27.3)	16 (72.7)	6 (27.3)								
Tibial	1	13 (59.1)	8 (36.4)	17 (77.3)	5 (22.7)	6.65	0.08	8.50	0.04	0.59	(0.42–0.75)	0.52	(0.31–0.71)
MNCV	2	13 (59.1)	9 (40.9)	13 (59.1)	9 (40.9)								
abnormality	3	17 (77.3)	5 (22.7)	18 (81.8)	4 (18.2)								
(Dr.'s judgment)	4	17 (77.3)	5 (22.7)	18 (81.8)	4 (18.2)								
Tibial	1	19 (86.4)	2 (9.1)	19 (86.4)	3 (13.6)	0.69	0.88	4.70	0.19	0.47	(0.15–0.79)	0.64	(0.32–0.93)
MNDL	2	19 (86.4)	3 (13.6)	20 (90.9)	2 (9.1)								
abnormality	3	20 (90.9)	2 (9.1)	21 (95.5)	1 (4.5)								
(Dr.'s judgment)	4	20 (90.9)	2 (9.1)	21 (95.5)	1 (4.5)								
Sural	1	9 (40.9)	13 (59.1)	7 (31.8)	15 (68.2)	12.78	<0.01	14.38	<0.01	0.60	(0.43–0.74)	0.56	(0.40–0.72)
SNAP	2	15 (68.2)	7 (31.8)	14 (63.6)	8 (36.4)								
abnormality	3	8 (36.4)	14 (63.6)	7 (31.8)	15 (68.2)								
(Dr.'s judgment)	4	11 (50.0)	11 (50.0)	11 (50.0)	11 (50.0)								
Sural	1	12 (54.5)	10 (45.5)	13 (59.1)	9 (40.9)	14.82	<0.01	17.40	<0.01	0.50	(0.32–0.66)	0.46	(0.29–0.63)
SNDL	2	17 (77.3)	5 (22.7)	18 (81.8)	4 (18.2)								
abnormality	3	8 (36.4)	14 (63.6)	8 (36.4)	14 (63.6)								
(Dr.'s judgment)	4	12 (54.5)	10 (45.5)	12 (54.5)	10 (45.5)								

**Wilcoxon signed-rank tests were not statistically significant between days 1 and 2.*

Shaded boxes represent statistically significant differences among neurophysiologist results. Shaded boxes indicate p values ≤0.05.

≤5th/≥95th percentile, interobserver differences among clinical neurophysiologists were found in 3 of 16 such comparisons. When the criterion was changed to ≤1st/≥99th percentile, no significant interobserver disagreement was found, but this latter comparison may be invalid because too few patients had this degree of abnormality to allow valid testing.

The agreement among investigators using the Krippendorff α was 0.67 ± 0.19. No intraobserver difference was observed between days 1 and 2 (footnote to Table 4).

Statistically Significant Interobserver Difference of Attributes of NC in Trials 3 and 4. Using the designation of D for significant interobserver differences for both days, AD for significant difference on only 1 day, and A (agreement) for no difference for either day, it was possible to compare the frequency of interobserver disagreement between Trials 3 and 4 (Table 5). Using measured raw values, agreement was observed in only 1 of 16 observations in Trial 3 and 6 of 16 observations in Trial 4. For normal deviates the frequencies were: 2 of 16 for Trial 3 and 6 of 16 for Trial 4. For NIS points from percentiles the ratios were 7 of 16 and 13 of 16, respectively. The data clearly imply improved interobserver agreement in Trial 4 when compared with Trial 3.

Clinical Neurophysiologist Judgment of Neurophysiologic Diagnosis of Polyneuropathy versus Judgment Based on Percentile Abnormality of Fibular MNCV and Sural SNAP Using a Standard Reference Source (RDNS-HS). Statistically significant disagreement was shown for 1 day for each criterion used to diagnose polyneuropathy

Table 4. Judgment of abnormality from use of defined percentile abnormality (≤2.5th/≥97.5th) and standard reference values* on attributes of nerve conduction: interobserver disagreement and agreement (and intraobserver disagreement†).

Nerve conduction attribute	Cl. NPhys	Number (%) of 24				Cochran Q test for differences among 4 Cl. NPhys teams				Krippendorff α (ordinal)			
		Day 1		Day 2		Day 1		Day 2		Day 1		Day 2	
		Normal	Abnormal	Normal	Abnormal	Q	*P*	Q	*P*	α	Bootstrap 95% CI	α	Bootstrap 95% CI
Fibular	1	16 (72.7)	4 (18.2)	17 (77.3)	5 (22.7)	7.60	0.06	3.00	0.39	0.49	(0.23–0.71)	0.71	(0.52–0.86)
CMAP	2	20 (90.9)	1 (4.5)	18 (81.8)	2 (9.1)								
abnormality	3	17 (77.3)	5 (22.7)	17 (77.3)	5 (22.7)								
(ND ≤2.5th)	4	20 (90.9)	2 (9.1)	19 (86.4)	3 (13.6)								
Fibular	1	13 (59.1)	5 (22.7)	15 (68.2)	4 (18.2)	5.23	0.16	3.00	0.39	0.78	(0.63– 0.90)	0.94	(0.85– 1.00)
MNCV	2	13 (59.1)	7 (31.8)	14 (63.6)	5 (22.7)								
abnormality	3	14 (63.6)	5 (22.7)	13 (59.1)	6 (27.3)								
(ND ≤2.5th)	4	17 (77.3)	4 (18.2)	14 (63.6)	6 (27.3)								
Fibular	1	18 (81.8)	2 (9.1)	19 (86.4)	1 (4.5)	2.40	0.49	4.71	0.19	0.54	(0.24–0.78)	0.68	(0.41–0.89)
MNDL	2	19 (86.4)	1 (4.5)	17 (77.3)	2 (9.1)								
abnormality	3	16 (72.7)	3 (13.6)	17 (77.3)	2 (9.1)								
(ND ≥97.5th)	4	18 (81.8)	3 (13.6)	17 (77.3)	3 (13.6)								
Tibial	1	19 (86.4)	2 (9.1)	18 (81.8)	2 (9.1)	3.00	0.39	–	–	0.90	(0.73–1.00)	1.00	(1.00–1.00)
CMAP	2	19 (86.4)	3 (13.6)	19 (86.4)	3 (13.6)								
abnormality	3	19 (86.4)	3 (13.6)	19 (86.4)	3 (13.6)								
(ND ≤2.5th)	4	19 (86.4)	3 (13.6)	19 (86.4)	3 (13.6)								
Tibial	1	14 (63.6)	6 (27.3)	15 (68.2)	4 (18.2)	1.84	0.61	2.61	0.46	0.43	(0.24–0.61)	0.36	(0.15–0.58)
MNCV	2	16 (72.7)	5 (22.7)	17 (77.3)	4 (18.2)								
abnormality	3	13 (59.1)	6 (27.3)	13 (59.1)	7 (31.8)								
(ND ≤2.5th)	4	13 (59.1)	8 (36.4)	14 (63.6)	7 (31.8)								
Tibial	1	18 (81.8)	2 (9.1)	16 (72.7)	3 (13.6)	8.59	0.04	11.33	0.01	0.35	(0.10–0.58)	0.58	(0.37–0.77)
MNDL	2	18 (81.8)	3 (13.6)	19 (86.4)	2 (9.1)								
abnormality	3	14 (63.6)	7 (31.8)	14 (63.6)	7 (31.8)								
(ND ≤97.5th)	4	14 (63.6)	7 (31.8)	15 (68.2)	6 (27.3)								
Sural	1	12 (54.5)	10 (45.5)	14 (63.6)	8 (36.4)	9.27	0.03	12.60	<0.01	0.65	(0.50– 0.79)	0.68	(0.54–0.81)
SNAP	2	16 (72.7)	6 (27.3)	17 (77.3)	5 (22.7)								
abnormality	3	11 (50.0)	11 (50.0)	11 (50.0)	11 (50.0)								
(ND ≤2.5th)	4	15 (68.2)	7 (31.8)	12 (54.5)	10 (45.5)								
Sural	1	14 (63.6)	3 (13.6)	15 (68.2)	2 (9.1)	3.67	0.30	–	–	0.74	(0.53–0.91)	0.87	(0.70–1.00)
SNDL	2	15 (68.2)	2 (9.1)	15 (68.2)	3 (13.6)								
abnormality	3	12 (54.5)	1 (4.5)	12 (54.5)	2 (9.1)								
(ND ≥97.5th)	4	12 (54.5)	4 (18.2)	11 (50.0)	2 (9.1)								

ND, normal deviate.

**Rochester Diabetic Neuropathy Healthy Subject Cohort (RDNS-HS) used to produce standard reference values.*
†Wilcoxon signed-rank tests were not statistically significant between days 1 and 2.

Shaded boxes represent statistically significant differences among neurophysiologist results. Shaded boxes indicate p values ≤0.05.

[i.e., clinical neurophysiologist judgment as compared with transformation of their measured values to percentiles and declaring abnormality when fibular MNCV and sural SNAP were ≤2.5th percentiles based on healthy subject reference values from the RDNS-HS). The degree of agreement among investigators was considerably higher using the latter approach (Table 6).

DISCUSSION

Attributes of NC are useful measurements for detection, characterization, and following the course of peripheral neuropathy. They are therefore useful in medical practice, epidemiologic surveys, and therapeutic trials.[3–12] By contrast, because they are physiological measures, they have less value for quantifying the clinical severity of muscle weakness or large-fiber sensory loss, and they typically have no value for characterizing or quantifying small sensory or autonomic fiber dysfunction. Therefore, attributes of NC are generally considered to be surrogate measures of polyneuropathy.[13]

Because attributes of NC are quantitative assessments that provide numeric values, it may be assumed that they are reliable indicators of abnormality and have no or low intra- or interrater variability. However, these assumptions are incorrect. Proficiency (accuracy and no or low intra- and interobserver variability) is heavily dependent on how NCs are performed, assessed, and interpreted. NCs can (and may be) performed and interpreted inaccurately.

Table 5. Statistically significant* interobserver differences for both days 1 and 2 (D), for only 1 of the 2 days (AD) and for neither day (A) in CI. NPhs Trials 3 and 4

	Trial					
	NCs measured units		NCs expressed as normal deviates† (from percentiles)		NCs expressed as NIS points (3 categories)‡	
Nerve conduction attribute	3	4	3	4	3	4
Fibular CMAP	D	D	D	D	A	A
Fibular MNCV	AD	AD	A	AD	A	A
Fibular MNDL	D	AD	D	AD	D	A
Tibial CMAP	D	D	D	D	D	A
Tibial MNCV	D	A	D	A	D	A
Tibial MNDL	D	D	D	D	AD	D
Sural SNAP	D	D	D	D	AD	AD
Sural SNDL	D	A	D	A	AD	A
McNemar S	3.57		2.67		3.60	
exact test *P*	0.1250		0.2188		0.1094	

**Using the Friedman χ^2 test.*
†Normal deviates-from percentile values of measured values corrected for applicable values (e.g., age, height, and others).
‡NIS points-3 percentile categories: ≤1st percentile = 2; >1st ≤5th percentile = 1; >5th percentile = 0. If the abnormality is in the other tail of the normal distribution, the order is reversed.

The question of NC proficiency has been tested directly in 2 previous trials utilizing expert clinical neurophysiologists.[2,14] Neither trial identified significant intraobserver differences in their NC assessments, but they both found significant interobserver differences. Also, some of these interobserver differences were of sufficient magnitude to make it of concern, especially for use in therapeutic trials. The degree of interobserver variability in the 2 trials was surprising, because, in the first, clinical neurophysiologists came from the same medical center and, in the second, investigators had been trained by physicians at the same center.

Trial 4 addressed 2 questions: (1) Could interobserver differences of measured NC attributes be reduced or eliminated by pretrial agreement to use exactly the same techniques of testing? (2) Could judgment of abnormality of individual attributes of NC and clinical neurophysiologic diagnosis of polyneuropathy be improved by use of common reference standards and setting abnormality by defined percentile levels of abnormality? Interobserver differences in assessed measured attributes of NC were less frequent in Trial 4 than in Trial 3. Presumably this decrease related to use of highly standardized techniques of NC assessment. However, one should note that interobserver disagreement was not eliminated. There were still differences in measured attributes of NC abnormality despite rigorous attempts to eliminate

Table 6. Clinical neurophysiologists' judgment and standard percentile/reference values for diagnosis of polyneuropathy: interobserver disagreement and agreement (and intraobserver disagreement*).

		Number (%) of 24				Cochran Q test for differences among 4 CI. NPhys teams				Krippendorff α (ordinal)			
		Day 1		Day 2		Day 1		Day 2		Day 1		Day 2	
Nerve conduction attribute	CI. NPhys	Normal	Abnormal	Normal	Abnormal	Q	*P*	Q	*P*	α	Bootstrap 95% CI	α	Bootstrap 95% CI
Neuropathy (Dr.'s judgment)	1	9 (40.9)	13 (59.1)	7 (31.8)	15 (68.2)	6.00	0.11	11.88	<0.01	0.73	(0.60–0.86)	0.63	(0.47–0.76)
	2	12 (54.5)	10 (45.5)	13 (59.1)	9 (40.9)								
	3	13 (59.1)	9 (40.9)	13 (59.1)	9 (40.9)								
	4	12 (54.5)	10 (45.5)	12 (54.5)	10 (45.5)								
Neuropathy (Σ fibular CMAP and sural SNAP ≤2.5th)	1	12 (54.5)	10 (45.5)	13 (59.1)	9 (40.9)	9.00	0.03	5.18	0.16	0.86	(0.76–0.96)	0.83	(0.71–0.94)
	2	15 (68.2)	7 (31.8)	15 (68.2)	7 (31.8)								
	3	12 (54.5)	10 (45.5)	12 (54.5)	10 (45.5)								
	4	12 (54.5)	10 (45.5)	13 (59.1)	9 (40.9)								

**Wilcoxon signed-rank tests were not statistically significant between days 1 and 2.*
Shaded boxes represent statistically significant differences among neurophysiologist's results. Shaded boxes indicate p values ≤0.05.

them. The fact that intraobserver differences were not found yet interobserver differences were found in all 3 trials suggests that the NC techniques are being performed reproducibly by individual clinical neurophysiologists. However, Trial 4 results also suggest that subtle differences of technique remain despite pre-instruction and agreement on techniques of test performance. Small differences in the EMG instrumentation used may conceivably account for some of the differences observed, but they probably would not explain difference of selective abnormality of some attributes of NC.

Are the observed interobserver differences of sufficient magnitude to be of concern for conduct of medical practice or therapeutic trials? For use in medical practice, the small differences observed do not appear to be meaningful clinically (see later), although accurate assessment of NC results with small interobserver differences remains a goal for conduct of both medical practice and therapeutic trials. For therapeutic trials, even very small interobserver differences are of concern, because they decrease the statistical power of a study and may affect the success of a trial. Therefore, our studies support the conclusions of earlier studies[2] that the most optimal therapeutic trial performance results when the same clinical neurophysiologists perform all serial NCs on a given subject.

This trial also demonstrates the value of using common reference values, setting abnormality, and the use of an agreed-upon percentile abnormal value.[15] Interobserver disagreement of judgment of abnormality of individual attributes of NC was reduced using common reference values and defining their abnormality by a given percentile as compared with individual clinical neurophysiologists' judgment. The latter judgments may relate to different reference values among medical centers. These data provide a rationale for use of common reference standards and defined percentile abnormality among EMG laboratories.

Could the choice of subjects explain the observed significant interobserver differences of measured attributes of NC in the Cl. NPhys 3 and 4 Trials? It seems unlikely that the somewhat advanced age of subjects would account for the demonstrated interobserver differences of measured attributes of NC. Conceivably it could have affected judgment of abnormality for individual attributes of NC assessed and for judgment of polyneuropathy. The patients' median age of 70 years required the electrophysiologists to distinguish changes related to "normal" aging from those due to diabetes. This proved challenging, presumably because of limited normative data available to investigators for subjects of advanced age and because the lower limit of normal for some attributes approaches the response detection limit. In this context, median values for sural SNAP amplitudes ranging from 2 to 5 μV represented significant intraobserver differences, but are unlikely to represent a physiologically important difference, emphasizing the use of broader categories of NC abnormality based on standard reference values corrected for applicable variables.

This trial (Cl. NPhys 4) is the part of a series of trials by our group aimed at assessing the proficiency of neuropathy signs (Trials 1 and 2), attributes of NC (Trial 3 and this trial), quantitative sensation testing (Trial 5), and perhaps others.[14,16,17]

Proficiency testing is now expected in the assessment of most laboratory tests. Such proficiency testing has not been required of clinical assessments of neuropathy symptoms and signs or of clinical neurophysiologic tests, presumably due to the impracticality. For laboratory proficiency testing, it is possible to send a small sample of serum or tissue to a reference or multiple reference laboratories, which can be done on multiple occasions and on a continuing basis. The same proficiency procedures cannot be followed for assessment of patients' examinations. However, it is possible, as shown in our Cl. NPhys trials, to do proficiency testing occasionally to assess critically the proficiency of physician clinical or neurophysiology test assessments.

APPENDIX

Additional Cl. NPhys Trial 4 Investigators. *Additional members of the Coordinating Committee:* P. James B. Dyck, MD, and Phillip A. Low, MD (Department of Neurology, Mayo Clinic, Rochester, MN, USA). *Study neurologists and diabetologists:* Henning Andersen, MD (Aarhus University Hospital, Aarhus, Denmark); John D. England, MD (Department of Neurology, Louisiana State University, New Orleans, LA, USA); Gareth Llewelyn, MD (University Hospital of Wales, Cardiff, Wales, UK); Michelle L. Mauermann, MD (Department of Neurology, Mayo Clinic, Rochester, MN); Dinesh Selvarajah, MD (Royal Hallamshire Hospital, Sheffield, UK); Wolfgang Singer, MD (Department of Neurology, Mayo Clinic, Rochester, MN); A. Gordon Smith, MD (University of Utah, Salt Lake City, UT); Solomon Tesfaye, MD (Royal Hallamshire Hospital, Sheffield, UK); and Adrian Vella, MD (Division of Endocrinology, Diabetes, Metabolism and Nutrition, Mayo Clinic, Rochester, MN, USA).

The authors thank Mary Lou Hunziker, Department of Neurology, Mayo Clinic, for preparation of the manuscript.

REFERENCES

1. Dyck PJ, Albers JW, Wolfe J, Bolton CF, Walsh N, Klein CJ, et *al.* A trial of proficiency of nerve conduction: greater standardization still needed. Muscle Nerve 2013;48:369–374.
2. Chaudhry V, Cornblath DR, Mellitis ED, Avila O, Freimer ML, Glass JD, et *al.* Inter- and intra-examiner reliability of nerve conduction measurements in normal subjects. Ann Neurol 1991;30:841–843.
3. Kimura J. Nerve conduction and needle electromyograpy. In: Dyck PJ, Thomas PK, editors. Peripheral neuropathy, 4th ed. Vol. 1. Philadelphia: Elsevier; 2005. p 899–970.
4. Thompson PD, Thomas PK. Clinical patterns of peripheral neuropathy. In: Dyck PJ, Thomas PK, editors. Peripheral neuropathy, 4th ed. Philadelphia: Elsevier; 2005. p 1137–1162.
5. Schifitto G, Yiannoutsos C, Simpson DM, Adornato BT, Singer EJ, Hollander H, et *al.* Long-term treatment with recombinant nerve growth factor for HIV-associated sensory neuropathy. Neurology 2001;57:1313–1316.
6. Dyck PJ, O'Brien PC, Oviatt KF, Dinapoli RP, Daube JR, Bartleson JD, et *al.* Prednisone improves chronic inflammatory demyelinating polyradiculoneuropathy more than no treatment. Ann Neurol 1982; 11:136–141.
7. Apfel SC, Schwartz S, Adornato BT, Freeman R, Biton V, Rendell M, et *al.* Efficacy and safety of recombinant human nerve growth factor in patients with diabetic polyneuropathy. A randomized controlled trial. JAMA 2000;284:2215–2221.
8. Ziegler D, Ametov A, Barinov A, Dyck PJ, Gurieva I, Low PA, et *al.* Oral treatment with alpha-lipoic acid improves symptomatic diabetic polyneuropathy: the SYDNEY 2 trial. Diabetes Care 2006;29:2365–2370.
9. Ziegler D, Low PA, Litchy WJ, Boulton AJ, Vinik AI, Freeman R, et *al.* Efficacy and safety of antioxidant treatment with alpha-lipoic acid over 4 years in diabetic polyneuropathy: the NATHAN 1 Trial. Diabetes Care 2011;34:2054–2060.
10. Berk JL, Suhr OB, Obici L, Sekijima Y, Zeldenrust SR, Yamashita T, et *al.* Repurposing diflunisal for familial amyloid polyneuropathy: a randomized clinical trial. JAMA 2013;310:2658–2667.
11. Smith BE, Windebank AJ, Dyck PJ. Nonmalignant inflammatory sensory polyganglionopathy. In: Dyck PJ, Thomas PK, editors. Peripheral neuropathy, 4th ed. Philadelphia: Elsevier Saunders; 2005. p 2309–2320.
12. Grant IA, Benstead TJ. Differential diagnosis of polyneuropathy. In: Dyck PJ, Thomas PK, editors. Peripheral neuropathy, 4th ed. Vol. 1. Philadelphia: Elsevier; 2005. p 1163–1180.
13. Suanprasert N, Berk JL, Benson MD, Dyck PJB, Klein CJ, Gollob JA, et al. Toward representative quantitation of neuropathy signs and tests in transthyretin amyloidosis. 2014 (in preparation).
14. Dyck PJ, Overland CJ, Low PA, Litchy WJ, Davies JL, Dyck PJB, et *al.* Signs and symptoms versus nerve conduction studies to diagnose diabetic sensorimotor polyneuropathy: Cl vs. NPhys Trial. Muscle Nerve 2010;42:157–164.
15. Dyck PJ, Litchy WJ, Lehman KA, Hokanson JL, Low PA, O'Brien PC. Variables influencing neuropathic endpoints: the Rochester Diabetic Neuropathy Study of healthy subjects (RDNS-HS). Neurology 1995; 45:1115–1121.
16. Dyck PJ, Overland CJ, Low PA, Litchy WJ, Davies JL, Dyck PJB, et *al.* "Unequivocally abnormal" vs. "usual" signs and symptoms for proficient diagnosis of diabetic polyneuropathy Cl vs. N Phys Trial. Arch Neurol 2012;69:1609–1614.
17. Dyck PJ, Argyros B, Russell JW, Gahnstrom L, Naleppa S, Albers JW, et al. Multicenter trial of the proficiency of smart quantitative sensation tests. Muscle Nerve 2013 Aug 8. [Epub ahead of print]. PMID: 23929701. DOI:10.1002/mus.23982.