MULTICENTER TRIAL OF THE PROFICIENCY OF SMART QUANTITATIVE SENSATION TESTS

PETER J. DYCK, MD,[1] BARBARA ARGYROS, R EEGT, R PhT, NCVT,[2] JAMES W. RUSSELL, MD,[2] LINDE E. GAHNSTROM, BA,[1] SUSAN NALEPA, R NCS.T, CNCT,[3] JAMES W. ALBERS, MD,[3] KAREN A. LODERMEIER, AA,[1] ANDREW J. ZAFFT, R EMG/EP T,[1] P. JAMES B. DYCK, MD,[1] CHRISTOPHER J. KLEIN, MD,[1] WILLIAM J. LITCHY, MD,[1] JENNY L. DAVIES, BA,[1] RICKEY E. CARTER, PhD,[4] L. JOSEPH MELTON III, MD,[5] and MEMBERS OF THE CL VERSUS NPHYS TRIALS (APPENDIX)

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

Accepted 24 July 2013

ABSTRACT: *Introduction*: We assessed proficiency (accuracy and intra- and intertest reproducibility) of smart quantitative sensation tests (smart QSTs) in subjects without and with diabetic sensorimotor polyneuropathy (DSPN). *Methods*: Technologists from 3 medical centers using different but identical QSTs independently assessed 6 modalities of sensation of the foot (or leg) twice in patients without ($n = 6$) and with ($n = 6$) DSPN using smart computer assisted QSTs. *Results*: Low rates of test abnormalities were observed in health and high rates in DSPN. Very high intraclass correlations were obtained between continuous measures of QSTs and neuropathy signs, symptoms, or nerve conductions (NCs). No significant intra- or intertest differences were observed. *Conclusions*: These results provide proof of concept that smart QSTs provide accurate assessment of sensation loss without intra- or intertest differences useful for multicenter trials. Smart technology makes possible efficient testing of body surface area sensation loss in symmetric length-dependent sensorimotor polyneuropathies.

Muscle Nerve **49**: 645–653, 2014

Loss of sensation is a common manifestation of various sensory and sensorimotor polyneuropathies.[1–5] These polyneuropathies are common and of diverse cause: metabolic alterations (diabetes mellitus, uremia, and others); infections (HIV, herpes, leprosy, syphilis, Lyme borreliosis, and others); malnutrition, vitamin deficiency, and alcoholism; intoxications (medicinal and industrial); inflammatory immune conditions; genetic causes and others. Detection, characterization and quantification of the kind, distribution, and severity of sensation loss and of heightened sensory phenomena (positive neuropathic sensory symptoms and tactile and thermal hyperalgesia) are useful for detection, differential diagnosis, and follow-up of these diseases.[6–8] Following the course of sensation loss over time may be needed in therapeutic trials and in monitoring effectiveness of therapy of patients on treatment regimens. Increasingly, quantitative sensation tests (QSTs) are being used in therapeutic trials, e.g., in diabetic sensorimotor polyneuropathy (DSPN), transthyretin

Abbreviations: ATTR-PN, transthyretin amyloid polyneuropathy; CASE IVb and c, a design by PJD and colleagues, Computer Assisted Sensory Examination – the version marketed by WR Medical Electronics, Maplewood, called Computer Aided Sensory Evaluator.; C disc, cooling discrimination using Dyck thermal disks; CDT, cooling detection threshold; DM, diabetes mellitus; DSPN, diabetic sensorimotor polyneuropathy; HP 0.5, heat-as-pain detection threshold; HP 5, an intermediate threshold level of heat-as-pain severity from 1–10; Nds, normal deviates from percentiles; NIS, Neuropathy Impairment Score; NSC, Neuropathy Symptoms and Change Score; QSTs, quantitative sensation tests; "smart" QSTs, smart quantitative sensation tests; TP DT, touch-pressure detection threshold; VDT, vibratory detection threshold

Key words: accuracy and reliability of nerve tests; diabetic sensorimotor polyneuropathy; intra- and intertest reproducibility; neurophysiology tests; smart quantitative sensation tests

The authors have full access to all of the data and the right to publish any and all data, separate and apart from the guidance of any sponsor. No other disclosure is reported by any author of this study.

Disclosures: P.J. Dyck serves as an Associate Editor for *Diabetes* and receives an honorarium. Neither he nor Mayo Clinic receives financial support from manufacturers of quantitative sensation testing equipment or peripherals. (i.e., WR Medical Electronics, Maplewood, MN and North Coast Medical, Inc., Morgan Hill, CA). He has made published testing approaches, algorithms of testing, and reference values corrected for applicable variables available to WR Medical Electronics, which facts are reported in CASE IVb test reports. B. Argyros reports no disclosures; J.W. Russell reports no disclosures; L.E. Gahnstrom reports no disclosures; S. Nalepa reports no disclosures; J.W. Albers reports no disclosures; K. Lodermeier reports no disclosures; A.J. Zafft reports no disclosures; P.J.B. Dyck reports no disclosures; C.J. Klein reports no disclosures; W.J. Litchy reports no disclosures; J.L. Davies reports no disclosures; R.E. Carter reports no disclosures; L.J. Melton, III reports no disclosures

Statistical analysis completed by: J.L. Davies, R.E. Carter, and P.J. Dyck

Authors' contributions: P.J. Dyck – study concept/design, acquisition of data, analysis/interpretation, critical writing and revision of the manuscript and study supervision; B. Argyros – examined patients and participated in the writing of the report; J.W. Russell – design and execution of the study and writing of the report; L.E. Gahnstrom – examined patients and participated in the writing of the report; S. Nalepa – examined patients and participated in the writing of the report; J.W. Albers – design and execution of the study and writing of the report; K. Lodermeier – examined patients and participated in the writing of the report; A.J. Zafft – examined patients and participated in the writing of the report; P.J.B. Dyck – design and execution of the study and writing of the report; C.J. Klein – design and execution of the study and writing of the report; W.J. Litchy – examined patients, design and execution of the study and writing of the report; J.L. Davies – design and execution of the study, analysis/interpretation and writing of the report; R.E. Carter – design and execution of the study, analysis/interpretation and writing of the report; L.J. Melton, III – design and execution of the study and writing of the report

Study funding: Research reported in this publication was supported in part by Mayo Foundation Funds and grants obtained from the National Institute of Neurological Disorders and Stroke (R01-NS36797), Dr. P. J. Dyck, PI and the National Institute on Aging (R01-AG34676), Dr. W. A. Rocca, PI. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

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

Published online 8 August 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/mus.23982

amyloid polyneuropathy (ATTR-PN), and Fabry disease.[6,9,10]

This study focuses on the performance of specific smart QSTs that are being used increasingly in therapeutic trials. These QSTs are computer controlled (smart), standardized, and referenced. They provide printed test results of what was done and found, and results are compared with reference values. They are tests of both large and small sensory fiber functions. Because of these design features, it is assumed that they can provide accurate information about sensory loss, which should be the same among participating medical centers and over time. However, this assumption, although reasonable, must actually be tested in field studies, which is the purpose of this trial. It might be argued that improved assessment of sensation using smart QSTs is not needed, because better assessments can be done by expert physicians. However, this assumption must now be questioned, because it has been shown that even evaluations by expert neuromuscular physicians were not as accurate and reproducible as had been assumed.[11,12] However, expert neuromuscular physician performance could be improved markedly by consensus development and the use of the simple criterion of judging only unequivocal abnormality while taking age, gender, and physical fitness into account.[11,12] Therefore, to obtain proficient physician examinations, considerable selection, training, and surveillance of physician performance is necessary. Therefore if some evaluations done by physicians could be performed by trained technologists using smart QSTs, therapeutic trials might be performed more efficiently and accurately.

In this trial (Cl vs. NPhys Trial 5), we assess not only the accuracy but also the intra- and intertest reproducibility of 6 highly standardized and referenced QSTs administered by technologists from 3 different medical centers in the same masked patients without and with DSPN.

RESEARCH DESIGN AND METHODS

Trial Context. Previous studies in this series of clinical versus neurophysiology tests have evaluated the proficiency of the clinical assessment of signs and symptoms (Trials 1 and 2)[11,12]; attributes of nerve conduction (Trials 3[13] and 4 [being prepared for publication]); and in this trial (Trial 5) of smart QSTs in DSPN. Considerable attention had been given at an earlier time to making all aspects of QST as standard, automated, referenced, and efficient as possible to limit the role of the technologist to ensuring highly standardized instruction and testing. This standardization was intended to obtain test result reproducibility among medical centers and over time.

Smart Quantitative Sensation Tests ("Smart" QSTs). By "smart" QST, we refer to use of special technologies (instruments), computer software, and standard test conditions to make QST as standard, reproducible, referenced, automated, and efficient as possible so that trained technologists, under the supervision of an expert physician, can assess cutaneous sensation accurately and reproducibly at predetermined anatomical sites as proficiently as possible. In "smart" QST, all aspects of sensation evaluation (i.e., instruction of subjects, stimuli used, algorithms of testing and finding threshold, comparison to reference values, and printout of what was done and observed) should be standard, quantitated, described in detail, and validated in the scientific literature. A final report should be generated which summarizes patient and evaluator information, all stimuli given, choices made, estimated threshold, and comparison to reference values.

In this trial, we tested both hand- and instrument-administered stimuli and both large and small sensory fiber functions. For hand-administered stimuli of large fiber function, we assessed touch-pressure threshold using 9 graded monofilaments and small fiber function using Dyck thermal disks. For instrument-administered stimuli of large fiber function we assessed vibration, and for small fiber function we assessed cooling and heat as pain (CASE IVc, WR Medical Electronics, Maplewood, MN) sensations.

Monofilament testing of touch pressure threshold was performed with Semmes Weinstein monofilaments that were modified by us so that monofilaments A, B, C - - - I produced static loads of -3, -2, -1 - - - 5 ln gms at 5/6 of their extended lengths.[8] Nineteen magnitudes of touch-pressure can be tested using this approach.[8] We used a slightly modified, previously described, algorithm of forced-choice 2:1 stepping to assess touch-pressure threshold.[8] The procedure of testing was demonstrated to technologists by a video presentation and also in a short training session lasting ~3 to 4 h before the formal trial described below. To provide highly standardized hand-held testing the technique of application of the monofilament is described here in detail and illustrated in Figure 1.

Application of the monofilament is at right angles to the surface of the skin. The tip of the monofilament is brought to within 1 or 2 mm of the skin surface, gently lowered to make skin contact, then depressed further to bend the monofilament to 5/6 of its extended length, then slowly released. The entire sequence, from contact to breaking of contact, should take ~1.5–2 s and be done smoothly. For null stimuli, the technologist performs a sham movement without actually making contact with the skin. Testing is done with the

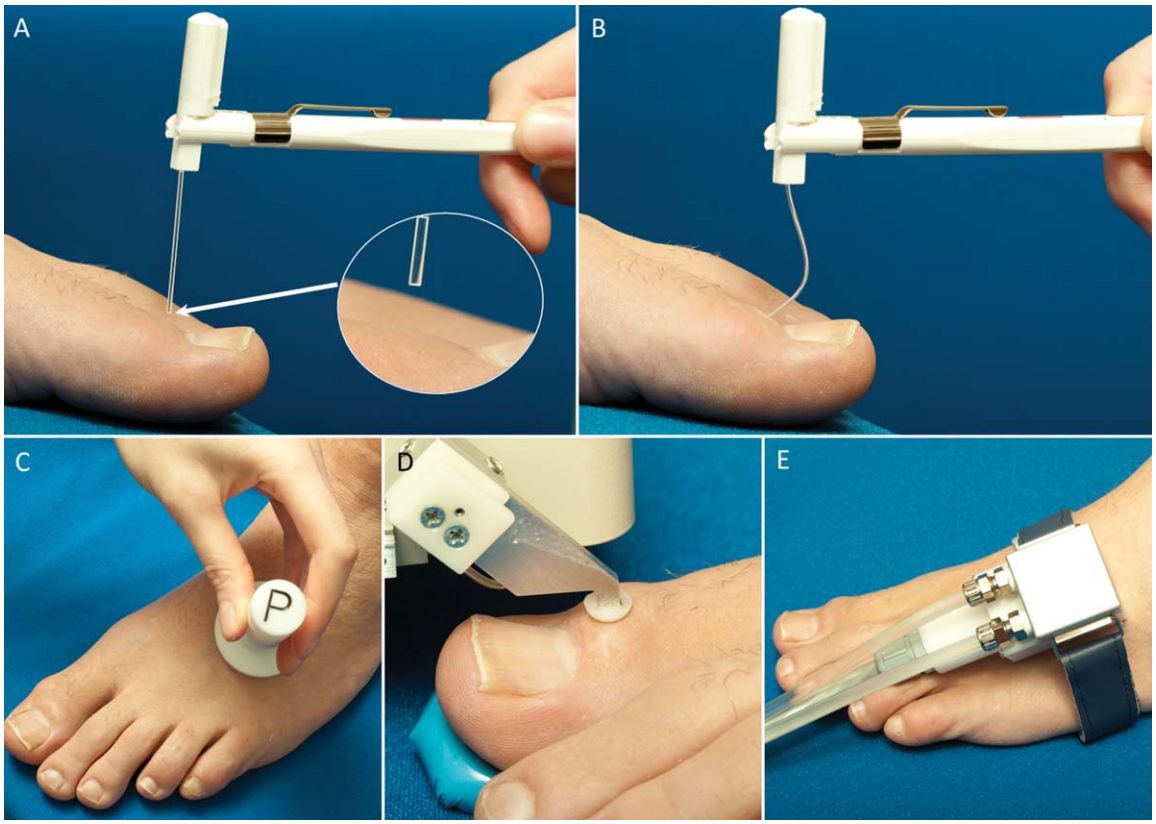

FIGURE 1. Features of Smart QST. **A,B:** Monofilament testing is illustrated. **C:** Testing with Dyck thermal disks is shown. **D:** In vibratory sensation testing with CASE IVc, vibratory stimuli are superimposed on a static load to allow eloquent and quantitative control of stimuli and null stimuli. **E:** The standard thermode in place on the foot to test cooling and heat-as-pain thresholds. Standard thermal pulses are superimposed on a small offset from skin temperature to assess modalities of thermal sensation. The restraining strap must be tightened just to the point at which the thermode makes full contact with the skin.

subject's eyes closed or with use of a blindfold if needed. The patient must say whether he/she was touched during periods 1 or 2 (the random sequence provided by the computer program); the examiner says "1" then "2" for the 2 intervals of forced-choice testing.

Cooling discrimination (C disc) was tested with Dyck Thermal Disks (described previously, Fig. 1).[14,15] Thermal disks have standard dimensions, weights, and appearances (except for the identification notation on the end of the handle not visible to subjects or patients. The different test thermodes have different heat transfer characteristics based on different material on their surfaces. A standard 2:1 forced-choice stepping algorithm was used to determine threshold with initial published reference values.[15]

The algorithm of testing and reference values corrected for anatomical sites and applicable variables was provided by us but programmed for personal computer and the CASE IV instrument (WR Medical Electronics, Maplewood, MN).

Vibratory detection threshold (VDT), CDT, and HP 0.5 and an intermediate severity of heat pain (HP 5) were evaluated using CASE IVc (initially developed by us[16] and later manufactured by WR Medical Electronics, Maplewood, MN but without our proprietary involvement; see author disclosure). The 4, 2, and 1 stepping algorithm with null stimuli was used to determine VDT and CDT.[7] For assessment of HP 5 (intensity 5 of 1–10), an ascending nonrepeating stepping algorithm with null stimuli was used.[7] The number of stimuli and null stimuli used in CASE IV testing has been described previously and is standard.[7] In forced-choice testing of touch pressure and cooling discrimination, the numbers of pairs of stimuli and null stimuli are variable depending on age, threshold of sensation, and performance of subjects.[8,15]

Typically, technologists could perform the 6 QSTs within the allotted time period of 1 h and 20 min. Approximately half of the time was spent in setting up the tests and providing instruction. Tests like vibratory and touch-pressure may be done in 3 to 6 min, whereas estimation of HP 0.5 and HP 5 may take a somewhat longer period of time.

Training of QST Technologists. Three technologists from different medical centers (Baltimore, MD [B.A.], Ann Arbor, MI [S.N.], and Rochester, MN [L.G.]) performed the QST. All had training and

experience with QST, but only 1 (L.G.) performed QST regularly. The 2 neurophysiology technologists who performed QST infrequently (B.A. and S.N.) were sent instructional material, a training video, sets of monofilaments and thermal disks, and the personal computer programs (provided by WR Medical Electronics, Maplewood, MN) for monofilament and thermal disk testing.

On the day preceding the 4-day trial at Rochester, MN, the standard testing procedures were reviewed for ~3 to 4 h, demonstrated, and practiced on healthy subjects. The emphases in this brief training session were: (1) use of standard assessments, (2) use of the same verbal instruction using cueing cards, (3) standard application of stimuli at standard anatomical sites, (4) need to keep the subject attentive, (5) correct performance of standardized testing and entering of responses, and (6) final printing out of test results.

Selection of Trial Subjects. Twelve subjects, 6 healthy and 6 persons with DSPN, were recruited for study from the Rochester Diabetic Neuropathy Study of healthy subjects[17] and patients with DM.[3] The neuropathy status was assessed after obtaining consent for the study by determining attributes of nerve conduction and performance of the Neuropathy Impairment Score. Subjects signed informed consent approved by the Mayo Clinic Institutional Review Board and were paid an honorarium to offset time away from other activities.

Conduct of CI Versus NPhys Proficiency Trial 5 (the Present Trial). To keep research costs low, the 2 technologists from Maryland and Michigan performed studies on 12 subjects at Rochester, MN, rather than have the 12 subjects travel to those medical centers. The studies were performed on December 6–9, 2011, in the QST Laboratory at Mayo Clinic, Rochester, MN. Arrangements were made to have completely independent QST assessments of the 12 subjects by each technologist on 2 occasions. Three different QST CASE IVc instruments and hand-held testing devices were used by the 3 technologists. Technologists were asked to not obtain subject or disease information during QST. The order of QST was randomized so that 3 subjects were evaluated concurrently for all 6 modalities of sensation by each of the 3 technologists in a half day period (54 QSTs). Approximately 80 minutes was made available for performance of the 6 QSTs. The tests were repeated, in random order, on the third or fourth day (a total of 432 QSTs). In each testing session, the order of tests followed by each technologist was: TP DT, C disc, VDT, CDT, and HP 0.5, and HP 5.0. Immediately after each QST session, automated printed results were handed to supervisory study personnel without revision or alteration.

Testing was done on the dorsum of the left foot of all subjects except for the heat pain test, which was done on contiguous regions of the lateral leg to avoid a change in threshold due to repeat testing and to avoid possible thermal injury. Monofilament testing was done on the dorsal terminal phalanges of all 5 toes. Vibratory testing was done on the dorsal surface of the terminal phalanx of the first toe. Cooling threshold was tested on the flat dorsal surface overlying the distal half of the metatarsal bones (Fig. 1).

Results of QSTs were compared with standard neurologic assessment of signs (Neuropathy Impairment Score [NIS]), symptoms (Neuropathy Symptoms and Change [NSC]) and a composite score of attributes of nerve conduction (Σ5 NC nds $\leq$ 2.5th percentile).[17–19]

Analysis. Standard statistical tests were used to assess accuracy and intra- and interobserver differences (see Tables (1–3)). The ICC in Table 2 was calculated using a two-way mixed effects model with each measurement obtained from 1 individual.[20] Because QST, NIS, NSC, and Σ5 NCs were measured by different scales, they were standardized before calculating the ICC.

RESULTS

The Frequency of QST Abnormalities. The raw values of QSTs and comparative neuropathy tests (NIS, NSC, and Σ5 NCs) values are shown in Table 1. None of the healthy subjects (1–6) had abnormality of signs (NIS), symptoms (NSC), or nerve conduction (Σ5 NCs) assessed by masked evaluation (Table 1). Of the 6 healthy subjects only subject 2 had a QST abnormality; the mean value of VDT was in the abnormal range (i.e., $\geq$ 97.5th percentile). Because VDT in patient 2 was abnormal in 4 of 6 individual evaluations and almost abnormal in the remaining 2, it is likely that this patient has an abnormality of VDT of the great toe (for reasons not determined) despite not having signs, symptoms, or NC abnormalities. None of the mean values of the other QST modalities were abnormal in this subject. A few additional QST abnormalities of individual tests were found in other healthy subjects and in the following frequencies: monofilament 0/36, CDT 3/36, cooling discrimination 0/36, HP5 0/36, and HP 0.5 2/36. For the entire healthy subject group, QST abnormalities were observed in 9 of 216 QSTs (i.e., in 4%), close to the 2.5% we had set as a rigorous level of test abnormalities. If patient 2 is considered to actually have a VDT abnormality of the tested great toe, the frequency of QST abnormalities in healthy subjects falls below the 2.5 percentage level.

Table 1. Raw Values of Quantitative Sensation Test (QST) Results (Mean and SD) and Signs (NIS), Symptoms (NSC) and Composite Scores of Nerve Conduction (Σ5 NC nds) for CI vs. NPhys Trial 5 Assessing Proficiency of QSTs.

		Patients											
		Healthy subjects ($n = 6$)						Patients with diabetic polyneuropathy ($n = 6$)					
Parameter		1	2	3	4	5	6	7	8	9	10	11	12
VDT (JND)	Mean	13.2	17.7	10.4	19.8	13.5	19.5	21.3	22.3	20.0	26.0	26.0	21.7
(shaded if > 97.5th)	SD	1.3	1.9	0.9	0.9	0.9	1.0	1.2	0.6	3.5	0.0	0.0	1.2
Monofilament (JND)	Mean	2.0	3.3	2.3	5.2	2.2	4.5	6.5	13.2	6.7	19.0	19.0	14.3
(shaded if > 97.5th)	SD	2.0	2.1	1.9	2.5	1.9	2.2	2.0	1.6	0.8	0.0	0.0	2.3
CDT (JND)	Mean	10.7	7.3	7.9	9.1	–	8.3	12.8	21.8	16.8	26.0	26.0	21.2
(shaded if > 97.5th)	SD	2.6	2.5	1.1	1.7	–	2.7	4.3	2.2	5.5	0.0	0.0	2.5
Cooling Discrimination (JND)	Mean	1.3	1.7	1.3	1.7	1.3	1.7	2.7	12.5	6.5	13.0	13.0	6.8
(shaded if > 97.5th)	SD	0.5	0.5	0.5	1.0	0.5	0.8	1.6	0.5	2.1	0.0	0.0	2.0
HP:5 (JND)	Mean	22.8	24.7	23.9	22.7	24.2	23.1	22.1	26.1	24.1	29.8	29.5	23.8
(shaded if > 97.5th)	SD	0.5	1.1	0.5	0.5	1.0	1.5	0.6	2.7	0.7	0.4	0.8	0.4
HP:0.5 (JND)	Mean	20.7	20.6	22.2	19.3	22.0	22.5	20.6	22.7	21.3	24.9	24.4	20.4
(shaded if > 97.5th)	SD	0.7	1.3	0.4	1.3	0.6	1.4	0.8	1.7	1.5	1.6	1.8	1.4
NIS (pts) (shaded if > 2 pts)	Raw	0	0	0	0	0	0	0	21	2	30	79	39
NSC (pts) (shaded if > 1 pts)	Raw	0	0	0	0	0	0	1	3	1	6	9	6
Σ5 NC nds* (shaded if < 2.5th)	Raw	3.5	3.1	1.7	1.2	−0.1	−1.8	−8.1	−9.3	−12.2	−12.9	−14.2	−15.1

Sensitivity and specificity (95% CIs) are: VDT, 91.7% (83.3%, 91.7%) and 83.3% (83.3%, 91.7%); Monofilament, 100.0% (100.0%, 100.0%) and 66.7% (66.7%, 75.0%); CDT, 75.0% (66.7%, 83.3%) and 66.7% (65.0%, 66.7%); Cooling Discrimination, 100.0% (100.0%, 100.0%) and 58.4% (50.0%, 58.4%); HP 5, 100.0% (100.0%, 100.0%) and 33.3% (33.3%, 50%); HP 0.5, 100.0% (83.3%, 100.0%) and 33.3% (16.7%, 50.0%). For sensitivity and specificity, Σ5 NC nds is used as an indication of abnormality. See text for an explanation of why sensitivities and specificities of HP 5 may be lower than other modalities of sensations tested, i.e. different anatomical sites tested and possible difference in vulnerability of sensory nerve fibers in DSPN.

**Σ5 NC nds are mean values of measurable values of standard normal deviates from percentiles corrected for applicable variables obtained from the RDNS-HS database of fibular nerve compound muscle action potential (fib CMAP), fibular motor nerve conduction velocity (fib MNCV), fibular distal latency (fib MNDL), tibial MNDL, and sural sensory nerve action potential (sural SNAP) multiplied by 5. Abnormality of Σ5 NC nds was set at < 2.5th percentile based on study of 330 healthy subjects in the RDNS-HS cohort with Σ5 NC nd values plotted on age.*

Of the 6 patients with DSPN (patients 7–12 in Table 1), mean values of QSTs were abnormal in the following frequencies: VDT 6/6; monofilament testing of touch-pressure 4/6, CDT 4/6, cooling discrimination 3/6, HP 5 2/6, and HP 0.5 2/6. The sensitivities and specificities of the different modalities of sensation are provided as a footnote to Table 1.

Accuracy of QST Results Compared with Signs (NIS), Symptoms (NSC), and Composite Nerve Conduction Score (Σ5 NC nds ≤2.5th Percentile). An important measure of accuracy of QSTs is the correlation of QST results with measured continuous measures of severity of signs, symptoms, and nerve conduction assessed independently (Table 2). High ICC values were obtained for all QSTs except for heat as pain 5 (HP 5) and HP 0.5 (threshold). Very high correlations were found for both hand- and instrument-administered stimuli and for modalities of vibration, touch-pressure, cooling, and cooling discrimination, and as compared to nerve conductions, neuropathy signs (NIS), or neuropathy symptoms (NSC).

Lower ICC values were obtained for heat as pain threshold (HP 0.5) or intensity (HP 5). The lower correlation of heat as pain measures should not be attributed to lower proficiency of the heat as pain tests, because results are compared incorrectly with

Table 2. ICC of Mean Ranks* of 6 QST Tests with Ranks* of Σ5 NC nds, NIS, and NSC.

Parameter		VDT (nd)	Monofilament (JND)	CDT (nd)	Cooling Discrimination (JND)	HP:5 (nd)	HP:0.5 (nd)
Σ5 NC nds	ICC (3,1)	0.83	0.88	0.74	0.85	0.49	0.41
	95 % CI	(0.54, 0.95)	(0.66, 0.97)	(0.36, 0.92)	(0.58, 0.96)	(−0.05, 0.83)	(−0.14, 0.80)
NIS (pts)	ICC (3,1)	0.85	0.9	0.88	0.93	0.7	0.5
	95 % CI	(0.59, 0.96)	(0.72, 0.97)	(0.66, 0.97)	(0.78, 0.98)	(0.28, 0.91)	(−0.04, 0.84)
NSC (pts)	ICC (3,1)	0.88	0.93	0.87	0.95	0.58	0.44
	95 % CI	(0.67, 0.97)	(0.80, 0.98)	(0.64, 0.96)	(0.84, 0.99)	(0.08, 0.87)	(−0.11, 0.81)

**Ranks are the original rank divided by the number of non-missing patients.*

Table 3. Intra- and Inter-Observer QST Agreement in CI vs. NPhys Trial 5.

		n (%) of 12				Cochran Q test for differences among 3 technologists					
		Day 1		Day 2		Day 1		Day 2			
Parameter	Technologist	Normal	Abnormal	Normal	Abnormal	Q	*p*	Q	*p*	κ	*p*
VDT (JND)	1	6(50)	6(50)	5(41.67)	7(58.33)	2.00	0.37	1.00	0.61	0.83	0.0017
	2	7(58.33)	5(41.67)	6(50)	6(50)					0.83	0.0017
	3	7(58.33)	5(41.67)	6(50)	6(50)					0.50	0.0395
Monofilament (JND)	1	8(66.67)	4(33.33)	7(58.33)	5(41.67)	-[§]	-	2.00	0.37	0.82	0.0019
	2	8(66.67)	4(33.33)	8(66.67)	4(33.33)					1.00	0.0003
	3	8(66.67)	4(33.33)	8(66.67)	4(33.33)					1.00	0.0003
CDT (JND)*	1	7(58.33)	5(41.67)	5(45.45)	6(54.55)	1.00	0.61	-	-	0.46	0.0608
	2	7(58.33)	5(41.67)	7(58.33)	5(41.67)					1.00	0.0003
	3	6(50)	6(50)	7(58.33)	5(41.67)					0.83	0.0017
Cooling Discrimination (JND)	1	8(66.67)	4(33.33)	9(75)	3(25)	2.00	0.37	2.00	0.37	0.80	0.0023
	2	9(75)	3(25)	8(66.67)	4(33.33)					0.80	0.0023
	3	9(75)	3(25)	9(75)	3(25)					1.00	0.0003
HP 5 (JND)	1	9(75)	3(25)	9(75)	3(25)	2.00	0.37	2.00	0.37	1.00	0.0003
	2	10(83.33)	2(16.67)	10(83.33)	2(16.67)					1.00	0.0003
	3	10(83.33)	2(16.67)	10(83.33)	2(16.67)					1.00	0.0003
HP 0.5 (JND)	1[†]	9(75)	3(25)	9(75)	3(25)	4.50	0.11	2.67	0.26	1.00	0.0003
	2	12(100)	0(0)	11(91.67)	1(8.33)					0.26	0.1448
	3	9(75)	3(25)	9(75)	3(25)					1.00	0.0003

**One patient had faulty marking of test site, therefore was not included.*

[†]0.5 was added to the count in each cell before kappa was calculated.

[§]Cochran Q test could not be computed.

a less affected body site and to large fiber dysfunction expressed to a greater degree at a distal site. Therefore the correlation of the heat as pain test has not been adequately tested in this trial.

Intra- and Interobserver Differences of QST Results. This information is provided in Table 3. Using the Kappa (K) coefficient, a high degree of intraobserver reproducibility was observed for modalities of QSTs for each of the 3 technologists. The median K coefficients for technologists 1, 2, and 3 were 0.83, 0.92, and 1.00–all very high and not significantly different among technologists. Also, no significant interobserver difference was observed for assessment of QST results (Table 3). The bootstrap sample mean is 0.84 with a 95% confidence interval of 0.71, 0.92.

Illustrating Body Distribution of Sensation Loss in Symmetrical Length Dependent Sensorimotor Polyneuropathy. For following course of weakness, sensation loss, or autonomic deficit in therapeutic trials, it may not be sufficient to score abnormality at single anatomical sites; instead, the distributed deficit may need to be tested. Use of too few examination sites might produce too large a floor (or ceiling) effect for use in therapeutic trials. Considering assessment of QSTs for therapeutic trials a preferred approach might be to assess distributed sensation loss over the body surface area. To do such an assessment thoroughly would require excessive testing time. However, when the sensation loss is known to be symmetrical and length-dependent, it is possible to arrive at the distribution of sensation loss by use of efficient algorithms of testing. Using a specially designed computer software program and with known threshold values of the different distributed anatomical sites it is possible to estimate the distribution of sensation loss in a reasonable time period. The details of the approach will be described elsewhere but are illustrated here for a patient with diabetic sensorimotor polyneuropathy (Fig. 2).

DISCUSSION

The kind, severity, and body surface distribution of sensation decrease or heightened sensory phenomena (e.g., positive neuropathic sensory symptoms of "asleep numbness, prickling or stabbing, burning, or deep aching pain" or tactile or thermal hyperalgesia) need to be assessed in clinical medicine. Disorders with these manifestations are common and are the cause of considerable morbidity and high health care cost. While the clinical neurologic examination is typically used to assess these sensory phenomena, it is generally recognized that such assessment is frequently not done or is not done well. It is difficult for physicians using only hand-administered stimuli (cotton wool, stick pins, vibration of a tuning fork, or displacement of a digit) to reliably recognize

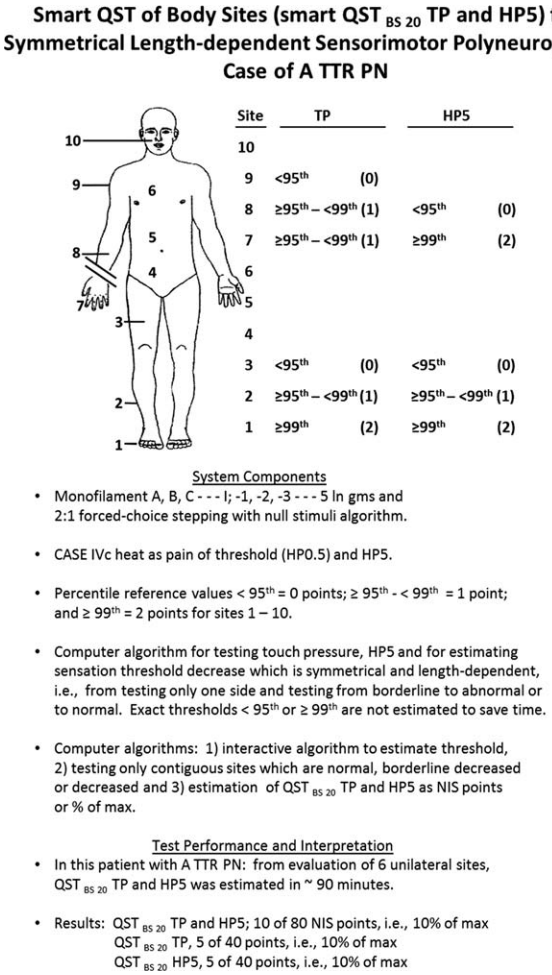

FIGURE 2. An illustration of smart QST of body surface areas to assess distributed sensation loss of touch pressure and heat as pain 5 (of 1–10) in a patient with transthyretin amyloid polyneuropathy.

thresholds of sensation that may vary with anatomical site, age, gender, and anthropomorphic variables. It is an even greater problem to assess these modalities of sensation over the surface of the body and over time. While it is recognized that physicians can achieve a high degree of proficiency using simple hand-held instruments, one cannot assume that physicians involved in therapeutic trials will manifest a sufficient level of proficiency to meet the high requirements of such trials. Usual sensory testing is not standardized sufficiently, referenced, or proficient to be used confidently as a primary outcome measure in therapeutic trials. In comparison it is now possible to use highly standardized QSTs that provide standardized and referenced smart QSTs by which to test sensation loss among medical centers and with time.

For QSTs to meet the need for use in therapeutic trials, stimuli should be appropriately chosen, should be of known waveform and magnitude, and stimulus magnitude should increase exponentially in steps from very small to very large to be suitable for neurosensory testing. Validated algorithms of testing and threshold finding should be used. Reference values corrected for applicable variables should be available for tested anatomical sites. For therapeutic trials, it would be desirable for all QST events to be programmed so that tests would be given and assessed by exactly the same procedures and be expressed with comparison to standard reference values corrected for applicable variables. In this trial, although both hand- and instrument-testing was used, all 6 QSTs fulfilled the standard criteria listed above, including instruction, hand-administered stimuli, and overall conduct of tests.

For multicenter trials of sensorimotor polyneuropathy, is assessment of sensation and its body surface distribution necessary? Would it suffice to assess only surrogate measures of sensation loss, e.g., attributes of nerve conduction of sensory fibers or of quantitative measures of epidermal nerve fibers? While this might suffice for some trials, for many other trials actual assessment of sensation loss is needed, because nerve conduction abnormality and epidermal nerve fiber counts may not represent the kind, severity, and distribution of sensation loss and have major floor (ceiling) effects. Also, the number of times that tissue can be biopsied is limited. Additionally, histologic tests are more detailed, time consuming, and expensive than QSTs.

This trial provides strong proof of concept that the highly standardized and referenced QSTs assessed here provided accurate information about touch-pressure, vibration, cooling, cooling discrimination, and heat as pain loss without intra- and intertest differences. The results suggest that if the tests were done in exactly the same way using the same reference values corrected for applicable variables, similar good results would be possible when done at other medical centers, assuming that technologist performance and patient responses were also the same. The good test performance observed here probably relates to: (1) the use of highly standardized stimuli and null stimuli, ranging from very small to very large in exponential steps; (2) use of validated algorithms of testing and finding threshold (which does not arrive at spurious results simply because insufficient replications are assessed); (3) use of reference values corrected for applicable variables of anatomical site, age, and anthropomorphic variables; and (4) handling of all aspects of testing, estimation of threshold, analysis of results, and printing out of results using computer technology. Obviously, the important role of technologists in performance of QST cannot be minimized. But use of the QST technology explained in these studies minimizes the tasks that technologists must perform to a minimum.

They provide instruction, observe patient alertness and cooperation, correctly administer the test, and perform related activities. It goes without saying that the QSTs should be administered only to cooperating patients.

Which modalities of sensation should be used for therapeutic trials or longitudinal follow-up of patients with sensorimotor polyneuropathy? Although this question was not addressed in this trial, several possibilities could be considered. It seems unlikely that users would want to include all of the QSTs studied here, especially knowing that several or multiple anatomical sites might need to be tested. Thus, for a trial of polyneuropathy in which large fibers are mainly involved (e.g., a trial of spinocerebellar degeneration), tests of touch-pressure or vibration might be suitable. In trials in which small fibers are affected preferentially (e.g., atypical small fiber polyneuropathy associated with diabetes mellitus or Fabry disease), heat as pain or cooling detection threshold might be tested. In transthyretin amyloid polyneuropathy (ATTR-PN) with involvement of both small and large sensory fibers, a measure of both large (touch-pressure) and small (heat as pain) sensory fiber involvement might be tested.

To make multiple anatomical site testing possible in a reasonable period of time it is important to use highly efficient algorithms of testing as illustrated in Figure 2. The possible time saving with the use of this algorithm comes from: (1) testing only 1 side (because the polyneuropathy is symmetrical); (2) beginning at a suitable length-dependent site, e.g., at lateral leg or volar forearm, and testing in both or either proximal or distal directions depending on response; (3) not taking the extra time needed to estimate threshold precisely, testing only to values $\geq$99th, $<$95th or $\geq$95th– $<$99th; (4) having previously determined the percentile levels of the 95th and 99th percentile for 10 anatomical sites; and (5) using computer software to manage all testing procedures, decision making, and printout of results so that technologists can reliably instruct, observe, hand-administer monofilament stimuli, and enter test responses.

The decision about which modality of sensation to use in multicenter trials might also depend on measures of sensitivity and specificity and correlation with surrogate measures of sensation loss. By these criteria, touch-pressure, vibration, and cooling thresholds performed at very high levels in the present trial. Heat as pain performed at a lower level, but its performance was not tested adequately. Large fiber sensory tests are compared appropriately to attributes of nerve conduction (both large fiber functions), but heat as pain (a small fiber function) should be compared with another measurement than nerve conduction (as done here).

The high degree of proficiency demonstrated in this trial was not unexpected, because previous trials of QST, in some cases using very similar technologies, provided strong evidence of test–retest reproducibility[3–5,21–27] and in a few trials of interobserver reproducibility also.[23,28]

What uses should be made of the proficient smart QSTs described here? While this question was not a focus of these studies, some possibilities come to mind. These approaches are being used for the study of metabolic and inherited polyneuropathy. Assessment of the course of sensation loss could be an important outcome measure in diabetic sensorimotor polyneuropathy. The QST $_{\text{BS 20}}$ TP and HP5 approach might reasonably be used for conduct of therapeutic trials of transthyretin amyloid polyneuropathy. For individual patients with sensorimotor polyneuropathy on a treatment trial in medical practice, QSTs evaluated here might be used to help provide information on efficacy of treatment.

In this trial, only loss of sensation was assessed without assessment of aberrant sensory phenomenon such as positive neuropathy sensory symptoms and measured tactile or thermal hyperalgesia. Although the latter 2 phenomena could be assessed by QST, they were not studied in this trial.

APPENDIX

Phillip A. Low, MD[1]; and Carol J. Overland[1] (Additional members of the Coordinating Committee) and Henning Andersen, MD[2]; John D. England, MD[3]; Gareth Llewelyn, MD[4]; Michelle L. Mauermann, MD[1]; Dinesh Selvarajah, MD[5]; Wolfgang Singer, MD[1]; Gordon Smith, MD[6]; Solomon Tesfaye, MD[7] and Adrian Vella, MD[8] (Study Neurologists and Diabetologists). From the [1]Department of Neurology, Mayo Clinic, Rochester, Minnesota; the [2]Department of Neurology, Aarhus University Hospital, Aarhus, Denmark; the [3]Department of Neurology, Louisiana State University, New Orleans, Louisiana; the [4]Department of Neurology, University Hospital of Wales, Cardiff, Wales, United Kingdom; the [5]Department of Human Metabolism, Sheffield Teaching Hospitals, Sheffield, United Kingdom; the [6]Department of Neurology, University of Utah, Salt Lake City, Utah; the [7]Department of Diabetes, Sheffield Teaching Hospitals, Sheffield, United Kingdom; and the [8]Division of Endocrinology, Mayo Clinic, Rochester, Minnesota.

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

REFERENCES

1. Dyck PJ, Kennel AJ, Magal IV, Kraybill EN. A Virginia kinship with hereditary sensory neuropathy: peroneal muscular atrophy and pes cavus. Mayo Clin Proc 1965;40:685–694.

2. Ohta M, Ellefson RD, Lambert EH, Dyck PJ. Hereditary sensory neuropathy, type II. Clinical, electrophysiologic, histologic, and biochemical studies of a Quebec kinship. Arch Neurol 1973;29:23–37.
3. Dyck PJ, Kratz KM, Lehman KA, Karnes JL, Melton LJ III, O'Brien PC, et al. The Rochester Diabetic Neuropathy Study: design, criteria for types of neuropathy, selection bias, and reproducibility of neuropathic tests. Neurology 1991;41:799–807.
4. Fagius J, Wahren LK. Variability of sensory threshold determination in clinical use. J Neurol Sci 1981;51:11–27.
5. Bertelsmann FW, Heimans JJ, Weber EJ, van der Veen EA, Schouten JA. Thermal discrimination thresholds in normal subjects and in patients with diabetic neuropathy. J Neurol Neurosurg Psychiatry 1985;48:686–690.
6. Goldberg JM, Lindblom U. Standardized method of determining vibratory perception thresholds for diagnosis and screening in neurological investigation. J Neurol Neurosurg Psychiatry 1979;42:793–803.
7. Dyck PJ, O'Brien PC, Johnson DM, Klein CJ, Dyck PJB. Quantitative sensation testing. In: Dyck PJ, Thomas PK, editors. Peripheral neuropathy, 4th ed. Philadelphia: Elsevier; 2005. p 1063–1094.
8. Dyck PJ, Winkler JA, Andrews KL, Kavros SJ, Vella A, Davies JL. Testing of touch-pressure sensation: introduction of the touch-pressure sensogram. In: Dyck PJ, Dyck PJB, Engelstad JK, Low PA, Amrami KK, Spinner RJ, Klein CJ, editors. Companion to peripheral neuropathy illustrated cases and new developments. Philadelphia: Saunders Elsevier; 2010. p 327–329.
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. Schiffmann R, Kopp JB, Austin HA III, Sabnis S, Moore DF, Weibel T, et al. Enzyme replacement therapy in Fabry disease: a randomized controlled trial. JAMA 2001;285:2743–2749.
11. 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.
12. 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.
13. Dyck PJ, Albers JW, Wolfe J, Bolton CF, Walsh N, Klein CJ, et al. A trial of proficiency of nerve conduction: greater standardization needed. Muscle Nerve 2012 [Epub ahead of print].
14. Dyck PJ, Curtis DJ, Bushek W, Offord K. Description of "Minnesota Thermal Disks" and normal values of cutaneous thermal discrimination in man. Neurology 1974;24:325–330.
15. Dyck PJ, Hansen JD, Winkler JA, Witt LV, Davies JL. Thermal disk quantitative sensation testing of cooling discrimination. In: Dyck PJ, Dyck PJB, Engelstad JK, Low PA, Amrami KK, Spinner RJ, Klein CJ, editors. Companion to peripheral neuropathy illustrated cases and new developments. Philadelphia: Saunders Elsevier; 2010. p 331–333.
16. Dyck PJ, Zimmerman IR, O'Brien PC, Ness A, Caskey PE, Karnes J, et al. Introduction of automated systems to evaluate touch-pressure, vibration, and thermal cutaneous sensation in man. Ann Neurol 1978;4:502–510.
17. 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.
18. O'Brien PC, Dyck PJ. Procedures for setting normal values. Neurology 1995;45:17–23.
19. Dyck PJ, O'Brien PC, Davies J, Klein CJ, Dyck PJB. Nerve tests expressed as percentiles, normal deviates, and composite scores. In: Dyck PJ, Thomas PK, editors. Peripheral neuropathy, 4th ed. Philadelphia: Elsevier; 2005. p 971–984.
20. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420–428.
21. Bravenboer B, van Dam PS, Hop J, vd Steenhoven J, Erkelens DW. Thermal threshold testing for the assessment of small fibre dysfunction: normal values and reproducibility. Diabet Med 1992;9:546–549.
22. Valensi P, Attali JR, Gagant S. Reproducibility of parameters for assessment of diabetic neuropathy. Diabet Med 1993;10:933–939.
23. Gelber DA, Pfeifer MA, Broadstone VL, Munster EW, Peterson M, Arezzo JC, et al. Components of variance for vibratory and thermal threshold testing in normal and diabetic subjects. J Diabetes Complications 1995;9:170–176.
24. Bril V, Kojic J, Ngo M, Clark K. Comparison of a neurothesiometer and vibration in measuring vibration perception thresholds and relationship to nerve conduction studies. Diabetes Care 1997;20:1360–1362.
25. Valk GD, de Sonnaville JJ, van Houtum WH, Heine RJ, van Eijk JT, Bouter LM, et al. The assessment of diabetic polyneuropathy in daily clinical practice: reproducibility and validity of Semmes Weinstein monofilaments examination and clinical neurological examination. Muscle Nerve 1997;20:116–118.
26. Chong PS, Cros DP. Technology literature review: quantitative sensory testing. Muscle Nerve 2004;29:734–747.
27. Moloney NA, Hall TM, O'Sullivan TC, Doody CM. Reliability of thermal quantitative sensory testing of the hand in a cohort of young, healthy adults. Muscle Nerve 2011;44:547–552.
28. Geber C, Klein T, Azad S, Birklein F, Gierthmuhlen J, Huge V, et al. Test-retest and interobserver reliability of quantitative sensory testing according to the protocol of the German Research Network on Neuropathic Pain (DFNS): a multi-centre study. Pain 2011;152:548–556.