A 4, 2, and 1 stepping algorithm for quick and accurate estimation of cutaneous sensation threshold

P.J. Dyck, MD; P.C. O'Brien, PhD; J.L. Kosanke, BS; D.A. Gillen, LPN; and J.L. Karnes, MSc

Article abstract—In quantitative sensory testing, certain methods may lead to incorrect estimates of vibratory (VDT), cool (CDT), or warm (WDT) detection thresholds. We have shown that the specific forced-choice algorithm of testing employed in our Computer-Assisted Sensory Examination (CASE IV) system, when compared with other tests of nerve dysfunction, provides accurate and reproducible estimates of these thresholds. Because this forced-choice algorithm is time consuming and performance might be made worse by drowsiness or boredom, we explored other algorithms that might provide estimates of threshold similar to those obtained with the forced-choice algorithm, but more quickly. In a trial of 25 healthy subjects and 25 patients with neuropathy, the 4, 2, and 1 stepping algorithm with null stimuli, based in part on comparative data from computer simulation and insights from patient decision making, provides an accurate estimate of threshold. On average, the time needed for forced-choice testing was 12.8 ± 2.9 minutes (mean ± SD). For 4, 2, and 1 stepping testing, it was 2.7 ± 2.5 minutes—a large saving of time. Since null stimuli were employed in the 4, 2, and 1 stepping algorithm, it was possible to monitor for spurious responses and repeat the test if they occurred at an excessive rate. The algorithm appears to be sufficiently robust to be recommended for clinical use and for some controlled clinical and epidemiologic trials.

NEUROLOGY 1993;43:1508-1512

Accurate and reliable assessment of touch-pressure, vibratory (VDT), cool (CDT), warm (WDT), or heat-pain (HPDT) detection thresholds depends not only on controlling subject and observer factors, testing environment, and stimulus characteristics, but also on the algorithm of testing and estimating thresh-

---

From the Peripheral Neuropathy Research Center, Department of Neurology (Dr. Dyck, and D.A. Gillen and J.L. Karnes), the Section of Biostatistics (Dr. O'Brien), and Statistical Systems (J.L. Kosanke), Mayo Clinic and Mayo Foundation, Rochester, MN.

Supported in part by grants from the National Institute of Neurological and Communicative Disorders and Stroke (14304) and from the Muscular Dystrophy Association.

Received September 18, 1992. Accepted for publication in final form December 22, 1992.

Address correspondence and reprint requests to Dr. Peter James Dyck, Mayo Clinic, 200 First Street S.W., Rochester, MN 55905.

old. Using the Computer-Assisted Sensory Examination system IV (CASE IV), designed and calibrated to present constant and quantitated stimuli over a broad range of stimulus intensities, we showed that the algorithm of testing and estimating threshold had a large influence on threshold. Estimating VDT with this system, using continuously increasing linear or exponential ramps of stimulus intensity from zero to the point of perception and signaled by depression of a response key (appearance threshold), resulted in a large overestimate of threshold when steep ramps were utilized.[1] This overestimate was attributed mainly to subject reaction times that were too slow, considering the rate of stimulus increase. We decided that algorithms employing ramps and appearance thresholds are inherently flawed because the response delay due to reaction time cannot be overcome and reaction times may vary, being greater in deliberate persons and in patients with neuromuscular disease. Use of very low-incline ramps, to minimize this error, would result in long testing times in insensitive parts or patients. Likewise, use of the mean of the appearance and disappearance threshold and exponentially increasing ramps to appearance threshold and decreasing ramps to disappearance thresholds also overestimated thresholds. The highest stimulus level was frequently exceeded before the response key was depressed, thus making it impossible to determine threshold at relatively insensitive sites. The shortcomings of algorithms employing continuously increasing ramps to appearance thresholds can be readily overcome by algorithms using stimuli that increase from baseline to a series of defined levels of intensity (25 levels in CASE IV), allowing a judgment to be made for each level (step) tested.[1-3] Since a judgment is made at each level tested before proceeding to the next level, the subject may take whatever time is needed to reach a decision or make a response. Slow responses do not spuriously raise threshold. Step stimuli are ideally suited for use in computer quantitative sensory testing because they permit use of stepping or forced-choice algorithms in which the level of stimulus intensity can be varied depending on response, and quick, complex algorithms of testing can be managed without error.

We previously have shown that use of a broad range of fixed stimulus intensities and a validated two-alternative forced-choice algorithm of testing provide reliable estimates of threshold.[4] The algorithm also has the meritorious feature of minimizing subject and observer bias. Since the specific rules of testing used in forced-choice algorithm may determine whether correct estimates of threshold are obtained, we validated our forced-choice algorithm by computer simulation and by testing associations of our estimates of sensory threshold with other measurements of sensory nerve function or structure.[4]

In this study, our objective was to develop a simple and valid stepping algorithm with null stimuli that would reproduce results obtained by our forced-choice algorithm but which would do so more quickly. If it was found to perform as well or almost as well as the forced-choice algorithm, it might find a place in medical practice and in epidemiologic studies. If a rapid but valid algorithm could be found, its use might be preferred over the forced-choice algorithm in certain test situations because prolonged testing, greater cost, boredom, and drowsiness could be avoided. The shorter test might allow assessment of more modalities of sensation or anatomic sites.

Methods. *The CASE IV system.* The design, components, and software of CASE IV have been described previously[2] and in the companion article.[5] In brief, the system may be used to assess VDT, CDT, WDT, or HPDT, using continuously increasing stimulus magnitudes to detection or using 25 fixed-intensity, pyramidal stimulus waves.

The initial 4, 2, and 1 stepping algorithms with null stimuli. The initial rules proposed were intuitive and based on previous experience. Testing was to begin at an intermediate level (level 13). The stimulus would be increased (if not felt) or decreased (if felt) by four steps to the point of turnaround (felt at the higher level when not felt at lower levels, or not felt at the lower level when it had been felt at the higher level). After the first turnaround, stepping was in steps of two. After the second turnaround, stepping was by steps of one. A total of 20 stimulus events were used, with five of them being randomly distributed null stimuli. If three consecutive failures were observed at level 25, testing was terminated, and the subject was classified as insensitive. If three consecutive successes were observed at level 1, testing was terminated and the subject was classified as supersensitive. Five null stimuli were randomly interspersed among 15 non-null stimuli. A positive response (indicating perception) to more than one null stimulus aborted the program. The patient was re-instructed, and the test was re-run. Three failures (due to spurious answers to null stimuli) when the test was re-run twice after the test was initially aborted indicated that the algorithm could not be used for this subject or patient.

Patient instruction. The following statement (in this case for VDT) was read to subjects and patients before the test:

> This is a test of your ability to feel vibration sensation. The test is not painful. It usually takes from three to four minutes. The stimulus may feel like vibration, buzzing, trembling, or rumbling. Some people cannot describe it, but they know a stimulus was given. All you have to do is decide whether you felt a stimulus during the display of the number "1." You probably will feel the stimulator resting on your toe (or finger) at all times. I will ask you to decide whether you felt an additional vibrating or other mechanical stimulus during number "1." After the number "1" has disappeared, you should push "yes" if you felt the vibration or some other mechanical stimulus, or "no" if you didn't feel it. Please get comfortable, relax your foot (or hand), and do your best. Once again, the object is to determine the smallest vibration or other mechanical stimulus you can feel. Do you have any questions?

Healthy subjects and patients. Twenty-five healthy volunteers, randomly selected from Rochester, MN, and

Table 1. The hypothetical responses chosen for use in computer simulation in patients assumed to have threshold (T) at step 13 (model A) or step 21 (model B)

Level	A (T = 13)	A05 (T = 13)	A10 (T = 13)	B (T = 21)	B05 (T = 21)	B10 (T = 21)
1	0.00	0.05	0.10	0.00	0.05	0.10
2						
3						
4						
5						
6						
7						
8						
9						
10	0.00	0.05	0.10			
11	0.10	0.145	0.19			
12	0.30	0.335	0.37			
13	0.50	0.525	0.55			
14	0.70	0.715	0.73			
15	0.90	0.905	0.91			
16	0.96	0.962	0.964			
17	0.96	0.962	0.964			
18	1.00	1.00	1.00	0.00	0.05	0.10
19				0.10	0.145	0.19
20				0.30	0.335	0.37
21				0.50	0.525	0.55
22				0.70	0.715	0.73
23				0.90	0.905	0.91
24				0.96	0.962	0.964
25	1.00	1.00	1.00	0.96	0.962	0.964

The table above indicates the probability of a positive response to challenges at each of 25 levels. Models A05 and B05 are modifications of models A and B, which assume that spurious positive responses occur in 5% of challenges in which the stimulus was not felt. Models A10 and B10 assume 10% spurious positive responses.

neurologically evaluated to exclude patients with neurologic disease and diseases known to predispose to neuropathy, had VDT, CDT, and WDT performed on the toe (VDT) and the dorsum of foot (CDT and WDT) using the initial 4, 2, and 1 stepping algorithm with null stimuli and also with the forced-choice algorithm.

An additional 25 patients, mostly with mild diabetic neuropathy, were evaluated by similar approaches.

Analysis of the data. The serial responses of each patient were inspected visually to identify whether the stepping algorithm was operating as anticipated or whether additional modification was needed. The serial responses were also examined to aid in identifying an appropriate method for estimating threshold. Insights gained from these inspections provided the basis for computer simulations evaluating ~~several stepping and~~ estimation algorithms. The results obtained in healthy subjects and neuropathy patients using the initial 4, 2, and 1 stepping algorithm with null stimuli were also compared with results using forced-choice testing in the same subjects.

The estimation algorithms considered included a general estimation algorithm that can be applied for any arbitrary pattern of testing (not just 4, 2, and 1 stepping) and an algorithm that consisted of taking the mean of the observed turnaround levels. In both cases, only levels arrived at by single stepping were considered in our analyses. In computing the mean of turnaround levels, if the final challenge did not result in a turnaround, a 21st challenge was added conceptually and assumed to result in a turnaround.

The regression estimator was obtained as follows. Let n represent the number of challenges that are arrived at by single stepping, let X_i be the level tested at the i^{th} such challenge, and let $Y_i = 1$ if a positive response is observed and 0 otherwise, $i = 1, \ldots, n$. Using least-squares regression, we estimate the quantities a and b in the equation $Y = a + bX$. The estimate of threshold is then given by the equation $T = (0.5 - a)/b$.

The simulation studies were carried out by assuming that a hypothetical patient would respond positively to a challenge at level i ($i = 1, \ldots, 25$) with probability denoted by P_i. Six such hypothetical models were considered, as shown in table 1. In model A, the assumption is that threshold occurs at level 13, so that $P_{13} = 0.5$. The probability of positive response rises to 1.0 by level 18 and declines to 0.0 by level 10. For such a hypothetical person, testing starts at level 13 by generating a random number between 0 and 1.0. If this number is less than 0.5, the subject is considered to have given a positive response. Otherwise, a negative response is assumed. After 15 such challenges, testing for that subject is terminated and threshold is estimated using each of the estimation algorithms under consideration. The process is then repeated (always using new random numbers) until 1,000 such hypothetical patients have been evaluated.

It became apparent from inspection of patient data that occasional spurious positive responses are obviously below the patient's actual threshold (table 2). To evaluate the effect of this phenomenon on estimates of threshold, we introduced the possibility of false-positive responses into the model. Thus, model A05 is similar to model A except that at each level, 5% of challenges that are not detected are assumed to result in false-positive responses (table 1). Model A10 increases this to 10%. Models B, B05, and B10 are similar except that the true threshold is increased to level 21 (table 1).

Results. The agreement between threshold estimated using the initial 4, 2, and 1 stepping algorithm (for both regression and mean of turnaround estimators) versus forced-choice testing is shown in the figure. It is seen that, with few exceptions, the agreement is good.

The 4, 2, and 1 stepping algorithm considerably shortened the time required for testing. After instruction, the time needed to find threshold averaged 2.2 ± 0.18 minutes (mean ± SD) for VDT; 2.2 ± 0.20 minutes for CDT; and 2.7 ± 0.76 minutes for WDT—times much shorter than from forced-choice testing.

Visual inspection of the pattern of responses observed in individual healthy subjects and neuropathy patients (figure) indicated that, in general, both the initial stepping and estimation algorithms performed well. However, there were some exceptions. If a subject reported feeling sensation to stimuli well below his or her actual threshold (see subject 1 in table 2), then multiple trials were needed to reach a level near the true threshold. In these instances in particular, but also in some others, we observed that the estimator of the mean of turnarounds gave more reasonable estimates of threshold than did the regression estimator.

Based on these insights, we simulated a variety of different stepping and estimation algorithms

Table 2. Examples in which the initial 4, 2, and 1 stepping algorithm performed poorly in healthy subjects and neuropathy patients

	Challenge* number															Threshold†
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
Neuropathy patient																
Patient 1																
Level	13	9	11	13	15	17	19	18	17	18	19	20	21	22	21	
Response‡	1	0	0	0	0	0	1	1	0	0	0	0	0	1	0	20.00
Patient 2																
Level	13	9	11	13	15	17	19	21	23	22	21	20	21	22	21	
Response	1	0	0	0	0	0	0	0	1	1	1	0	0	1	0	21.00
Patient 3																
Level	13	17	21	19	20	21	22	23	24	25	25	24	23	24	23	
Response	0	0	1	0	0	0	0	0	0	0	1	1	0	1	0	23.75
Healthy subject																
Subject 1																
Level	13	9	5	7	9	11	13	15	17	16	15	16	17	16	15	
Response	1	1	0	0	0	0	0	0	1	1	0	0	1	1	0	15.67
Subject 2																
Level	13	9	11	13	12	13	14	15	16	17	18	19	18	17	18	
Response	1	0	0	1	0	0	0	0	0	0	0	1	1	0	0	16.75

* Positions of null stimuli are not shown.
† Mean level based on single stepping.
‡ 0 = not felt, 1 = felt.

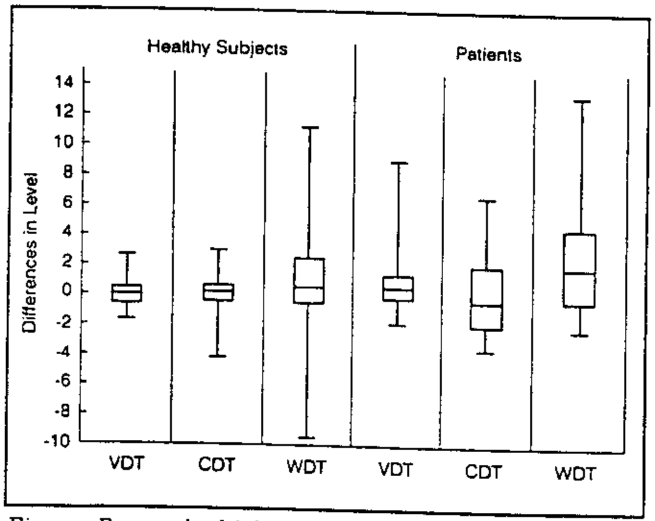

Figure. Box and whiskers figures to show the 25th, 50th, 75th, and range of differences in threshold (by level) between the 4, 2, and 1 stepping algorithm and the forced-choice algorithm (normalized to 0 level) for vibratory (VDT), cool (CDT), and warm (WDT) detection thresholds of the foot in 25 healthy subjects and 25 patients with neuropathy. Mean threshold levels were not significantly different between the two algorithms. Especially for healthy subjects, the difference in threshold for VDT and CDT was within one step of each other in 76% (VDT) and 68% (CDT) of subjects. In 92% (VDT) and 88% (CDT) of cases, it was not different by more than ± two levels. The variability was somewhat larger in patients with neuropathy, especially for WDT. These results suggest that the 4, 2, and 1 stepping algorithm with null stimuli performs well in reproducing the results of the forced-choice algorithm.

(table 3). The only modification to the stepping algorithm that was found helpful was to validate any positive response occurring to the first non-null challenge (at level 13). This was accomplished by repeating the challenge when a positive response was observed at the initial challenge, with the result of the second challenge used to determine whether the next test occurred at level 9 or level 17. No modifications to the estimation algorithm were identified that improved estimation relative to the mean of turnarounds approach.

In the final algorithm, we retest level 13 if a positive response has been given to the first challenge at level 13. Additionally, one of the null stimuli is randomly given during stimulus interval 1 or 2 (by random choice). The final change made to the initial algorithm was to use the mean of the turnaround levels arrived at by single stepping (for example, in the sequence 21, 19, 20, and 19, the first turnaround is at level 20).

The algorithm of testing and finding threshold was programmed into the computer so that automatic testing and printout of results were available.

Discussion. In the practice of both ophthalmology and audiology, psychophysical tests are being extensively used.[6] In the practice of neurology, psychophysical studies of sensation are not used extensively. In part, this has been due to (1) the perception that they were not needed, (2) the unavailability of good sensory testing systems, and (3) the unavailability of good algorithms of testing. With the availability of good systems to evaluate cutaneous sensation and the recognition that new infor-

Table 3. Comparison of different estimation-of-threshold algorithms using computer simulations

	Model A			Model A05			Model A10			Model B			Model B05			Model B10		
Method:	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
$\hat{T}$ percentiles																		
Min	9.3	11.1	10.8	-1.0	11.0	10.4	-24.0	7.3	8.0	-20.0	19.1	19.1	-20.0	12.0	16.4	$-\infty$	8.5	12.5
1	11.3	11.5	11.6	11.1	11.3	11.4	9.7	10.8	10.7	19.1	19.5	19.5	18.8	17.6	19.0	15.0	15.5	17.6
10	12.2	12.2	12.3	11.9	12.0	12.0	11.7	11.7	11.8	20.1	20.1	20.1	20.0	19.9	20.0	19.6	19.3	19.5
25	12.5	12.6	12.6	12.4	12.4	12.5	12.2	12.2	12.3	20.5	20.5	20.5	20.3	20.3	20.4	20.2	20.0	20.0
50	13.1	13.0	13.0	12.9	13.0	13.0	12.8	12.8	12.8	21.0	21.0	21.0	20.9	20.8	20.8	20.8	20.6	20.6
75	13.6	13.6	13.7	13.5	13.4	13.5	13.3	13.3	13.4	21.5	21.6	21.6	21.5	21.4	21.4	21.4	21.2	21.2
90	14.0	14.0	14.0	13.9	13.9	13.9	13.8	13.8	13.8	22.0	22.0	22.0	22.0	21.8	21.8	22.1	21.7	21.7
99	15.0	14.6	14.6	15.1	14.6	14.5	15.8	14.4	14.5	22.9	22.7	22.7	26.3	22.6	22.6	29.7	22.5	22.5
Max	21.3	15.7	15.7	21.0	15.7	15.7	37.5	15.7	15.7	38.0	23.3	23.3	64.0	23.3	23.3	74.3	23.3	23.3
% w/in ± 1 level	84.8	88.6	86.6	82.0	87.1	86.5	76.8	81.1	81.1	84.1	88.5	88.5	80.0	84.1	85.5	71.4	76.1	78.1

Method 1: 4-2-1 stepping, using results from all 15 challenges, T estimated by regression.
Method 2: 4-2-1 stepping, using only results from levels arrived at by single stepping, T estimated by mean of TAs.
Method 3: Same as method 2, but an S (positive response) at initial challenge is validated.

mation is provided by these approaches, it is anticipated that use of quantitative sensory testing will become more common.

Psychophysical approaches to human sensation in health and disease have lagged behind electrophysiologic study of sensory nerve function. Electrophysiologic studies of sensory nerve fibers have become important assessments of patients with sensory loss because the results produced are reproducible, objective (unaffected by the will of the subject), and accurate. They may provide information about whether dysfunction is in peripheral nerve or central tracts. Increasingly, however, there has been a renewed interest in the psychophysical study of sensation because such studies are needed (1) to assess sensory experience, (2) to discriminate function not separable in electrophysiologic studies, (3) to provide an overall assessment of sensation that combines the functions of cutaneous receptors, fiber pathways, and central processing, and (4) to provide information about cutaneous hypersensitivity phenomena. Assessment of threshold and suprathreshold responses are being used in controlled clinical trials, epidemiologic studies, and medical practice.

One need that the present study addresses is for a simple method of testing and estimating threshold that can be used for computerized sensory systems, that will provide accurate sensory thresholds, and that will do so quickly. The present 4, 2, and 1 stepping algorithm with null stimuli has been validated using both human subjects and computer simulation. We have shown that thresholds of VDT, CDT, and WDT are generally reproducible and usually not significantly different from those produced by forced-choice testing. Because the time needed to get threshold with the 4, 2, and 1 stepping algorithm is approximately one-quarter of the time required to get threshold with forced-choice testing, it provides a real saving of time. We believe the algorithm is sufficiently robust to be used not only for clinical practice but also for some epidemiologic and perhaps controlled clinical trials. Because thresholds were not shown to be significantly different from those in forced-choice testing, normative percentiles acquired from forced-choice testing may be used to approximate percentiles for 4, 2, and 1 stepping algorithm testing. The cost reduction due to the lesser time spent can be used either to reduce the charge for the test or to get more information, eg, testing several sites, to show a gradient of sensory abnormality in peripheral neuropathy.

References

1. Dyck PJ, Karnes JL, Gillen DA, O'Brien PC, Zimmerman IR, Johnson DM. Comparison of algorithms of testing for use in automated evaluation of sensation. Neurology 1990;40:1607-1613.
2. Dyck PJ, Karnes J, O'Brien PC, Zimmerman IR. Detection thresholds of cutaneous sensation in humans. In: Dyck PJ, Thomas PK, Lambert EH, Bunge R, eds. Peripheral neuropathy. 2nd ed. Philadelphia: WB Saunders, 1984:1103-1138.
3. O'Brien PC, Dyck PJ, Kosanke JL. A computer evaluation of quantitative algorithms for measuring detection thresholds of cutaneous sensation. In: Munsat TL, ed. Quantification of neurologic deficit. Boston: Butterworths, 1989:197-206.
4. 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;10:432-440.
5. Dyck PJ, Zimmerman I, Gillen DA, Johnson D, Karnes JL, O'Brien PC. Cool, warm, and heat-pain detection thresholds: testing methods and inferences about anatomic distribution of receptors. Neurology 1993;43:1500-1508.
6. Higgins HE, Jaffe MJ, Coletta NJ, Caruso RC, de Monasterio FM. Spatial constrast sensitivity; importance of controlling the patient's visibility criterion. Arch Ophthalmol 1984;102:1035-1041.