British Journal of Industrial Medicine 1988;**45**:635–639

Reliability of a widely used test of peripheral cutaneous vibration sensitivity and a comparison of two testing protocols

F E GERR, R LETZ

From the Division of Environmental and Occupational Medicine, Mount Sinai School of Medicine, New York, NY 10029, USA

ABSTRACT Quantitative, non-invasive instruments for assessing peripheral sensory function are being used in many epidemiological investigations of workplace hazards. To be useful in this context such tests should have high reliability and short administration time. The reliability and time efficiency of two testing protocols for determining cutaneous vibration sensitivity (vibration threshold) were compared in 22 healthy volunteers. Both methods were administered using a widely used testing device (Vibratron II). The first testing protocol was a two alternative "forced choice" method recommended by the instrument manufacturer. The second protocol was a "yes-no" method of limits procedure. Each subject's dominant index finger was tested with both methods on two occasions to compare their reliability. In these well educated subjects the method of limits procedure was found to be substantially more reliable and was much less time consuming than the recommended forced choice procedure. The simpler method of limits procedure may be preferable to the forced choice method in certain test conditions.

Several portable instruments for assessing peripheral sensory function have recently been introduced for use in occupational and environmental research,[1-3] are reviewed by Bove *et al.*[4] These instruments, and their associated methodologies, have many characteristics that render them useful as screening devices for peripheral sensory function abnormality. They allow for non-invasive, non-aversive, quantitative, and relatively rapid assessment of sensory function at multiple anatomical sites. They do not require a highly skilled technician to operate and testing may be readily performed in the field.

One of these instruments, the Vibratron II, is being used currently in several large epidemiological studies in the United States. The Vibratron II allows for the quantification of cutaneous vibration sensitivity of the fingers and toes. Initially intended as a screening device, it has recently gained United States Federal Drug Administration approval as a diagnostic device. Given the widespread use of the Vibratron II, it is remarkable that no study of the reliability of the manufacturer's recommended testing protocol has been published.

Accepted 24 August 1987

The Vibratron manufacturer recommends a two alternative forced choice method. Procedures of this type are considered insensitive to changes in the subject's criterion for the detection of a stimulus.[4] Methods of threshold determination that do not use a forced choice response, such as those requiring the subject to respond as to whether a stimulus is present or absent ("yes-no"), are potentially sensitive to subject criterion changes. For this reason, many investigators prefer the forced choice testing algorithms. The presumption is that the criterion changes inherent in yes-no methods result in a substantial increase in threshold variability and hence poor test reliability. Forced choice methods, however, are subject to considerable estimation error resulting in substantial threshold variability.[5,6] In fact, estimation error is considerably greater for forced choice procedures than for comparable yes-no procedures.[5,6]

In addition to reliability, the rapidity of administration is important in determining the usefulness of a testing protocol for epidemiological studies. It is optimal to test all four extremities in studies of peripheral neuropathy. This provides the basis for distinguishing between diffuse and focal neuropathic processes. Forced choice procedures can be time consuming to administer. A yes-no procedure, often

referred to as the method of limits, can be rapidly administered. Efficient methods allow for the testing of more individuals in a limited period.

In the current investigation we tested healthy volunteers using the Vibratron II and the two psychophysical procedures, the forced choice and method of limits. Both methods were performed on the same subjects on two occasions to compare their reliability and time efficiency.

Methods

The Vibratron II cutaneous vibration threshold testing instrument (Sensortek, Inc, Clifton, NJ) was used for all measurements. The instrument consists of a controller and two identical "slave" transducers. A plastic post, 1·5 cm in diameter, protrudes from each transducer. The transducer, via the post, provides a vibrating stimulus at 120 Hz. The amplitude of vibration is determined by the applied voltage, which is adjusted manually on the controller unit. The amplitude of vibration is quantified in "vibration units," digitally displayed on the controller unit. A small wooden platform was made to allow the hand and forearm to be in a relaxed position while the fingertip rests on the post.

Individuals were tested with the Sensortek forced choice protocol and with a method of limits protocol. Both methods (described below) were used to determine a vibration threshold for the index finger of the subject's dominant hand. The method of limits was also used to determine the vibration threshold for the index finger of the non-dominant hand and both great toes. All participants were tested on two separate occasions, between one and 40 days apart. All data were entered by the examiner directly from the digital display via keyboard into a COMPAQ microcomputer (COMPAQ Computer Corporation, Houston, Texas). The examiner did not have access during the second test session to data from the first session. In addition to recording the raw data, the computer automatically recorded the amount of time required to administer each threshold determination for both methods. Subjects were unable to view the control unit or the computer screen during testing.

The Sensortek staircase, two alternative forced choice procedure (SFC), is detailed in the operating manual for the Vibratron II. It requires the subject to determine which of the two posts is actually vibrating during any given trial. After each correct determination, the intensity of the vibration is decreased by 10% and the trial is repeated. A pseudorandom sequence is used to determine which of the two posts will provide the stimulus for each trial. The subject is required to guess when he or she is uncertain which of the two posts is actually vibrating. After the first incorrect response, the algorithm requires that two out of three determinations be correct before the stimulus intensity is lowered. Should two out of three determinations be incorrect, the stimulus intensity is raised. In addition, beginning at 0·7 vibration units, the two out of three criterion is used even if no errors have been made previously. Testing is terminated when five errors are made. A threshold is calculated by pooling the vibration settings of the five errors with the five lowest correct scores. The highest and lowest scores are dropped and the mean of the remaining eight values determine the threshold. In the current study an initial setting of 3·0 vibration units was used for all subjects, and if that setting was not clearly felt the setting was doubled until it was.

The method of limits yes-no procedure (MOL) began with delivery of suprathreshold stimulation that the subject could easily detect. The intensity of the stimulation was gradually reduced at a constant rate. The subject was asked to report verbally the earliest point in time at which he or she could no longer perceive the vibration. The setting was recorded, and the subject was asked to lift his or her finger from the stimulator post. The intensity of stimulation was reduced to well below the threshold of the previous trial and the subject was asked to place his or her finger back on the stimulator post. The intensity of the stimulus was gradually increased and the subjects were asked to indicate verbally when they could first feel the vibration in their finger. The complete testing sequence consisted of nine trials (five descending and four ascending). The terms method of limits and yes-no are both used in reference to the above described procedure. Specifically, however, the term method of limits refers to the determination of the delivered stimulus intensity and the term yes-no refers to the subject's response to the stimulus.

All participants provided informed consent. A self administered questionnaire was completed after the testing sessions. Demographic information included age, sex, race, current and previous work, and level of educational completion. Questions to determine the presence or absence of symptoms suggestive of neuropathy (numbness, tingling, weakness, and pain) or medical conditions known to predispose to sensory abnormalities (carpal tunnel syndrome, neck or back disorders, injury to the nerves of the arms or legs, alcoholism, diabetes, and kidney failure) were included. Prior exposures to neurotoxins were documented. A list of medications was obtained.

Twenty two subjects (12 male, 10 female) were tested on two occasions. There were 18 white, two black, and two hispanic subjects. The mean age was 35·4 years and ranged from 22 to 56. All subjects had completed high school and 20 had at least one advanced degree. None of the subjects had a history of

serious medical illness. Only one individual took medication on a daily basis (cholestyramine). One subject had been exposed to a potential neurotoxin (solvents and polyester resins).

Data from all participants were included in the analysis regardless of medical history or prior neurotoxic exposure. Test retest correlation coefficients (Pearson) and paired *t* tests between sessions were calculated for both SFC and MOL vibration thresholds. For the MOL procedure, the initial descending trial was discarded and vibration thresholds were calculated by averaging various subsets of the remaining eight trials. The test retest correlations of these various summary measures provided a criterion for comparing these alternative test summary measures.

Results

Table 1 presents comparisons of the results from the SFC and MOL procedures. Vibration thresholds determined by the SFC procedure were substantially lower than those determined by the MOL. The correlation of vibration thresholds between the two sessions, an index of the method's reliability, was much higher for the MOL than the SFC procedure. The SFC procedure took, on average, about twice as long to administer as the MOL procedure. For the MOL procedure the average vibration threshold was slightly lower in the second session when compared with the first. This suggests some systematic criterion shift between occasions, perhaps related to learning about how to respond. The difference in mean vibration threshold between sessions for the SFC procedure was negligible. Interestingly, the correlations between the SFC threshold and the corresponding MOL threshold were 0·55 and 0·59 respectively (not shown in table 1), the same magnitude as the correlation between occasions for the SFC procedure.

An effort was made to determine the cause of the relatively low correlation of SFC thresholds between the two sessions. Individuals' pairs of thresholds and records of the SFC trials generated during the determination of the thresholds, were examined. Although many pairs of thresholds were quite close, several individual records showed surprising discrepancies. Table 2 shows the actual SFC trials from both sessions for one such subject. As may be seen, on one of the two occasions correct choices resulted in diminution of the stimulus until zero (no vibration) was reached. The subject then responded correctly at zero four consecutive times. The SFC vibration threshold for that occasion was 0·0. The SFC threshold for the alternate occasion was 0·5. Other less dramatic examples were found in the data.

Another objective of this study was to investigate methods of deriving a vibration threshold from an average of various MOL trials or runs. The desired method would be time efficient and have high reliability. Values from various derivations of the MOL thresholds for the dominant index finger are presented in table 3. All summary measures, even the mean of the first two ascending runs, showed good reliability ($r > 0·7$). Thresholds with the highest reliability were obtained by averaging MOL runs 2–7 or by dropping the highest and lowest values from runs 2–5 ("trimmed 2–5"). Thus nine runs are not necessary

Table 1 *Comparison of method of limits (MOL) and Sensortek forced choice (SFC) procedures for estimating vibration thresholds for dominant index finger*

	MOL Mean (SD)	*SFC Mean (SD)*
Threshold session 1	1·60 (0·48)	0·82 (0·37)
Threshold session 2	1·46 (0·36)	0·82 (0·28)
Difference (session 1—session 2)	0·14 (0·29)	0·00 (0·33)
Correlation between sessions 1 and 2	0·81	0·51
Test duration session 1 (s)	233·4 (72·8)	442·3 (91·1)
Test duration session 2 (s)	185·3 (37·7)	370·7 (81·6)

Table 2 *Actual SFC trials from both sessions for a selected subject*

Session 1 *MOL estimate* = 1·3 *SFC estimate* = 0·0		*Session 2* *MOL estimate* = 1·2 *SFC estimate* = 0·5	
Stimulus	*Response*	*Stimulus*	*Response*
30	Correct	30	Correct
27	Correct	27	Correct
24	Correct	24	Correct
22	Correct	22	Correct
20	Correct	20	Correct
18	Correct	18	Correct
16	Correct	16	Correct
14	Correct	14	Correct
13	Correct	13	Correct
12	Correct	12	Correct
11	Correct	11	Correct
10	Correct	10	Correct
9	Correct	9	Correct
8	Correct	8	Incorrect
7	Correct	8	Correct
6	Incorrect	8	Correct
6	Correct	7	Incorrect
6	Correct	7	Correct
5	Correct	7	Correct
5	Correct	6	Incorrect
4	Correct	6	Incorrect
4	Correct	7	Correct
3	Correct	7	Correct
3	Correct	6	Correct
2	Correct	6	Correct
2	Incorrect	5	Correct
2	Correct	5	Correct
1	Correct	4	Incorrect
1	Incorrect		
1	Correct		
0	Correct		
0	Correct		
0	Correct		
0	Correct		
0	Correct		
0	Incorrect		
0	Incorrect		

Table 3 *Comparison of various summary measures for estimating vibration thresholds for the dominant index finger using the method of limits*

Average of runs	*Session 1 Mean (SD)*	*Session 2 Mean (SD)*	*Difference Mean (SD)*	*Correlation Pearson r*
2 to 9	1·60 (0·48)	1·46 (0·36)	0·14* (0·29)	0·807
2 to 7	1·62 (0·48)	1·47 (0·36)	0·15* (0·27)	0·829
2 to 5	1·64 (0·48)	1·48 (0·37)	0·16* (0·28)	0·816
2, 4, 6, 8	1·69 (0·51)	1·56 (0·35)	0·13* (0·33)	0·771
2, 4, 6	1·70 (0·50)	1·56 (0·35)	0·14* (0·31)	0·797
2, 4	1·70 (0·48)	1·56 (0·37)	0·14* (0·32)	0·742
Trimmed 2–9	1·61 (0·48)	1·47 (0·35)	0·14* (0·29)	0·809
Trimmed 2–7	1·65 (0·49)	1·49 (0·35)	0·16* (0·28)	0·815
Trimmed 2–5	1·66 (0·48)	1·51 (0·36)	0·15* (0·27)	0·828

*p < 0·05 (paired *t* test) that the mean difference is equal to zero.

and reliability with as few as five runs is quite good. A similar pattern was evident for the non-dominant finger and the toes. It should be noted that all the summary measures examined were, on average, lower for the second session when compared with the first. This suggests an improvement in performance of the test on the part of the subjects, or a systematic change in criterion for determining the presence or absence of the stimulus.

Table 4 presents summary statistics for a comparison of vibration thresholds of fingers and toes. These thresholds were derived by dropping the highest and lowest values from MOL runs 2–5 and averaging the remaining two runs. Note that the correlation between tests on two occasions (an index of reliability), high for all four sites, is higher for the toes than for the fingers. The latter effect is an artifact of the greater total variance in thresholds for the toes in comparison with the fingers (SDs for session 1 and 2). In fact, the absolute error in MOL thresholds (SD for difference between sessions) is higher for toes than fingers.

Discussion

In the population tested in this study the MOL procedure used was superior to the forced choice procedure recommended by the instrument manufacturer. The MOL procedure was substantially more reliable than the SFC and required about half as much time to administer. Investigation of alternative threshold estimate calculations for the MOL showed that the reliability from as few as five trials (trimmed 2–5) was essentially equal to that of threshold estimates determined from greater numbers of trials. This suggests that the administration time could be reduced by one third from the times reported here without adversely affecting the reliability of the MOL threshold estimate.

The current study is limited in its generalisability by at least two factors. Firstly, the study population had a low prevalence, if any, of neurological disease manifesting as peripheral neuropathy. Secondly, the level of educational attainment of the study population was high. The reliability of these methods in working populations with a lower level of educational attainment and at higher risk of peripheral neuropathy may vary from the reliability reported in this group of subjects. In addition, present work in our laboratory with culturally diverse groups and subjects with clinically overt neurological disease suggests that it is not always possible to obtain reliable MOL threshold estimates. It may be easier for a subset of these individuals to perform a forced choice procedure.

Several recent studies have compared the reliability of forced choice procedures for estimating vibration thresholds with more conventional non-forced choice methods. Bove *et al* (unpublished data) compared a forced choice best PEST[7] method administered with the Vibratron, to a method of limits procedure administered with an older instrument, the Biothesiometer.[8] The method of limits procedure was found to be more reliable than the forced choice procedure. The study did not compare the two

Table 4 *Estimated vibration thresholds using the method of limits for dominant and non-dominant fingers and toes*

Average of runs	*Session 1 Mean (SD)*	*Session 2 Mean (SD)*	*Difference Mean (SD)*	*Correlation Pearson r*
Dominant finger	1·66 (0·48)	1·51 (0·36)	0·15 (0·27)*	0·828
Non-dominant finger	1·60 (0·50)	1·37 (0·38)	0·23 (0·33)*	0·738
Dominant toe	4·16 (2·77)	3·68 (2·11)	0·48 (0·92)*	0·966
Non-dominant toe	3·70 (2·38)	3·69 (2·27)	0·01 (1·02)	0·905
Difference between hands	0·06 (0·25)	0·13 (0·28)*		
Difference between feet	0·46 (0·91)*	0·01 (0·59)		

*p < 0·05 (paired *t* test) that the mean difference is equal to zero.

procedures on the same instrument as did the current study. The Biothesiometer, however, has many shortcomings when compared with the Vibratron instrument. These include a diminution of vibration amplitude with increasing application pressure and the hand held nature of the instrument. The demonstration of superior reliability using the MOL with the Biothesiometer when compared with the Best PEST procedure administered with the Vibratron is probably due to differences in the testing procedure rather than to instrument differences.

Muijser *et al* compared a forced choice procedure with a method of limits using the Optacon tactile tester.[9] They concluded: "The use of the forced choice method in the present investigation was intended to reduce the effects of criterion setting on signal detection, thus reducing variability. However, unexpectedly the forced choice method did not produce lower standard deviations than did the much simpler method of limits" (p 295). Several subjects in this study were unable to provide valid thresholds when tested with the method of limits. These subjects failed, on runs of diminishing stimulus intensity, to indicate extinction of the perception of vibration. Because of this difficulty, the authors recommended the use of forced choice methods.

Kershaw has compared the properties of threshold estimates from forced choice and yes-no (method of limits) procedures both algebraically and by computer simulation.[5] He showed that large biases in threshold estimates can occur with forced choice procedures and concluded: "It appears that although a forced choice protocol is easy to operate, there are considerable and often hidden difficulties in obtaining estimates. Clearly some attempt should be made to replace it by some other procedure" (p 42).

The forced choice procedure recommended by the Vibratron manufacturer has difficulties in addition to those caused by estimation error inherent in two alternative forced choice procedures. It is difficult to interpret what location on the psychophysical curve the SFC is estimating. The SFC appears to be related to the up-down transformed response rules proposed by Wetherill and Levitt.[10] The Wetherill and Levitt rules estimate specific points on the psychophysical curve. They were not designed, however, to be used in forced choice experiments. In fact, when used with forced choice procedures such rules will cause levels to "drift toward low stimulus levels for which there is little or no chance of detecting the stimulus" (p 36).[5]

The rules for changing stimulus intensity in the SFC procedure should estimate the 0·50 point on the psychophysical curve in yes-no experiments; it is unclear what point is estimated in forced choice experiments. It does not estimate the 0·75 point and does not recover when strings of randomly correct guesses drive the stimulus intensity well below threshold (as evidenced by table 2). If a forced choice procedure is to be used with the Vibratron, or other similar instruments, the estimated threshold should be well defined with respect to the psychophysical curve. Investigators must be prepared to administer a substantial number of trials (100 or more) to achieve forced choice threshold estimates with reliability suitable for screening, diagnostic, or longitudinal research purposes.

The results of this study, and others, suggest that the current widespread uncritical acceptance of forced choice procedures as preferable to the method of limits may be unwarranted. Further elucidation of the variables that affect vibration threshold estimate reliability is necessary for all methods. Such variables might include severity of disease and level of educational attainment. The development of forced choice procedures that are less prone to estimation error is also needed. We recommend a method of limits procedure when the severity of disease in the population being tested is mild to moderate and when the subjects are fully able to understand and comply with the instructions. Finally, severe time constraints in some cases may compel investigators to use the MOL regardless of other considerations.

This study was supported by a grant from the US National Institute for Occupational Safety and Health (KO1-0H00064) and National Institute of Environmental Health Sciences Center Grant (P30-ES00928).

References

1 Arezzo JC, Laudadio C, Schaumburg HH. The Optacon tactile tester and the Pfizer thermal tester. New devices for the detection of diabetic neuropathy. *Diabetes* 1984; **33**(suppl):185A.

2 Arezzo JC, Schaumburg HH, Laudadio C. The vibratron: a simple device for quantitative evaluation of tactile vibration sense. *Ann Neurol* 1985;**35**(suppl 1):169.

3 Katims JJ, Naviasky EH, Ng Lky, Rendell M, Bleecher ML. New screening device for assessment of peripheral neuropathy. *J Occup Med* 1986;**28**:1219–21.

4 Bove F, Litwak MS, Arezzo JC, Baker EL. Quantitative sensory testing in occupational medicine. *Seminars in Occupational Medicine* 1986;**1**:185–9.

5 Kershaw CD. Statistical properties of staircase estimates from two interval forced choice experiments. *British Journal of Mathematical and Statistical Psychology* 1985;**38**:35–43.

6 Rose RM, Teller DY, Rendlemen P. Statistical properties of staircase estimates. *Perception and Psychophysics* 1970;**8**:199–204.

7 Lieberman HR, Pentland AP. Microcomputer-based estimation of psychophysical thresholds: the best PEST. *Behavior Research Methods and Instrumentation* 1982;**14**:21–25.

8 Goldberg JM, Lindbloom U. Standardized method of determining vibratory perception thresholds for diagnosis and screening in neurological investigation. *J Neurol Neurosurg Psychiatry* 1979;**42**:793–803.

9 Muijser H, Hooisma J, Hoogendijk EMG, Twisk DAM. Vibration sensitivity as a parameter for detecting peripheral neuropathy I. Results in healthy workers. *Int Arch Occup Environ Health* 1986;**58**:287–99.

10 Wetherill GB, Levitt H. Sequential estimation of points on a psychometric function. *British Journal of Mathematical and Statistical Psychology* 1965;**18**:1–10.