OBJECTIVE — Several national and international scoring systems are used to diagnose diabetic polyneuropathy (PNP). The variety in these scores and the lack of data on validity and predictive value has led to a comparison and validation of the scores with clinical standards for PNP to determine the most powerful measurement for screening.

RESEARCH DESIGN AND METHODS — Three matched groups were selected: 24 diabetic patients with neuropathic foot ulcers, 24 diabetic patients without PNP or ulcers, and 21 control subjects without diabetes. In all participants the scores from the International Consensus on the Diabetic Foot (ICDF) and the Dutch Nederlandse Diabetes Federatie-Centraal Beleids Orgaan (NDF/CBO) were tested. The Diabetic Neuropathy Symptom score, the Diabetic Neuropathy Examination score, Heart Rate Variability, the Nerve Conduction Sum score, and a San Antonio Consensus sum score were obtained as clinical standards. Reproducibility was tested in a separate study (13 patients).

RESULTS — The construct validity and discriminative power of the ICDF and NDF/CBO scores were comparable, although monofilaments (NDF/CBO) scored lower. The predictive value was good for all scores, with the best results being obtained for the tuning fork (NDF/CBO). Reproducibility of the NDF/CBO scores (monofilaments and tuning fork) was high.

CONCLUSIONS — The characteristics of the scores of tests recommended by ICDF and NDF/CBO are comparable. The single use of the 128-Hz tuning fork produces results similar to the extended scores of the ICDF and much better than those of monofilaments on validation and for predictive value. For screening we therefore advise the use of the tuning fork alone.

Diabetes Care 28:2201–2205, 2005

One of the most frequent complications of diabetes is polyneuropathy (PNP), a major cause to foot ulcers and amputations (1). The risk of amputation is a life-long threat to the diabetic patient, and the costs attributed to diabetic foot ulcers and amputations are high (2). Frequent assessment of risk factors is necessary for early detection of patients at risk, to prevent amputations due to diabetic foot disease (3). To select an adequate test for diagnosing diabetic PNP, validity, predictive value, and manageability are important criteria, as defined by Jaeschke et al. (4).

Several consensus statements give advice about diagnosing diabetic PNP. The San Antonio Consensus report states that at least one measurement should be performed in five different diagnostic categories (5). Various tests are available for all five categories, and most of them have been well validated. The use of all categories together leads to a large degree of overdiagnosis (6). Manageability in the outpatient clinic is difficult because of the large number and complexity of the tests that have to be performed.

The International Consensus on the Diabetic Foot (ICDF) advises that a set of tests should be performed to diagnose diabetic PNP (7). Two versions of this set are published in the same consensus report. This diagnostic set has not been validated yet, and the predictive value is uncertain. The manageability of this combination seems good.

The Dutch Nederlandse Diabetes Federatie-Centraal Beleids Orgaan (NDF/CBO) guidelines also advise diagnosing diabetic PNP by the combined use of tests, and different combinations are described in the same report (8). These sets of tests have not been validated, and the predictive value is uncertain.

Therefore, for none of the sets of tests mentioned are all of the criteria of Jaeschke et al. (4) met. We examined the value of the diagnostic tests of the ICDF and the NDF/CBO guidelines as tools for screening for diabetic PNP. The aim of this study is to compare these commonly used diagnostic tests for diabetic PNP with the purpose of defining a manageable and valid instrument with high predictive value to diagnose diabetic PNP in clinical practice.

RESEARCH DESIGN AND METHODS — Three groups of subjects were studied. The patients were recruited, after informed consent, from the Diabetes Outpatient Clinic (University Hospital Groningen). Patient records were screened consecutively during our...
outpatient clinics for patients with characteristics as described below.

Subjects with a history of or clinically apparent cardiac disease or electrocardiographic abnormalities or those using β-blockers or calcium antagonists were excluded. Subjects with peripheral arterial disease were excluded by normal ankle-arm indexes (>0.90), toe-arm indexes (>0.70), and normal plethysmography results (crest time 0.22 s) in all groups. Normal glucose tolerance of the control subjects was demonstrated by a fasting capillary blood glucose value <6.1 mmol/l and a blood glucose value <7.8 mmol/l 2 h after a 75-g oral glucose tolerance test. All groups were matched for sex and age (within 5 years), and the diabetic groups were matched for duration and type of diabetes (type 1/type 2) as well. Type 1 diabetes was defined using conventional criteria (on clinical grounds with BMI <27 and insulin dependence from onset of diabetes).

After this screening and selection subjects received appointments in a randomized order. The first group consisted of 24 diabetic patients with a history of neuropathic foot ulcers (diabetic ulcer [DU] group). These ulcers were purely neuropathic by origin, as was confirmed by their localization (plantar surface of the foot at high-pressure points) and the absence of peripheral arterial disease as described below. A second group (diabetes control subjects [DC] group) of 24 diabetic patients had no history of foot ulcers or clinically overt signs of neuropathy (defined as a complete lack of complaints or symptoms suggestive for PNP and a normal Achilles tendon reflex [ATR] as determined by the treating physician). The normal ATR is considered for this purpose as an early and reliable sign of the absence of PNP (9,10). The third group consisted of 21 control subjects with normal glucose tolerance (control group). Details of the clinical characteristics of each group are given in Table 1.

Tests from the ICDF and the NDF/CBO guidelines were performed in all subjects. As clinical standards, the following tests were obtained: the Diabetic Neuropathy Symptom (DNS) score, the Diabetic Neuropathy Examination (DNE) score, the Heart Rate Variability (HRV) as a test for cardiovascular autonomic neuropathy, the Nerve Conduction Sum (NCS) score as a electrodiagnostic set of tests, and an overall San Antonio Consensus (SAC) sum score.

Different researchers, blinded for the group to which the participant was allocated, performed the tests. The researchers were acting independently, and no information about the results was exchanged during the study.

ICDF
Two versions of tests are described to diagnose diabetic PNP (7): 1) ICDF 1 includes 128-Hz tuning fork, pin-prick testing at the hallux, Semmes-Weinstein monofilament (SW-MF) testing at the plantar surface of the foot, and ATR testing (7a); and 2) ICDF 2 includes 128-Hz tuning fork, pin-prick testing at the hallux, SW-MF testing at the plantar surface of the foot, and ATR testing combined with cotton wool testing at the hallux (7b). Items in both versions are scored from 0 to 2. A normal score is 0 points, a mild/moderate deficit is 1 point, and severely disturbed/absent response is 2 points.

NDF/CBO guidelines
Although a combination of four tests is advised (NDF/CBO 1), for daily clinical practice the use of SW-MF testing and/or 128-Hz tuning fork testing at the hallux is suggested, which leads to three more possible combinations (NDF/CBO 2–4) (8). 1) NDF/CBO 1 includes 128-Hz tuning fork and cotton wool testing at the hallux, SW-MF testing at the plantar surface of the foot (at the hallux and centrally at the heel), and ATR testing. Items in these tests are scored from 0 to 2. (Normal score is 0 points, a mild/mild/moderate deficit is 1 point, and severely disturbed/absent response is 2 points.) 2) NDF/CBO 2 is tuning fork testing. A 128-Hz tuning fork is used to examine vibration perception at the dorsum of the interphalangeal joint of the right hallux. The vibrating tuning fork is put on the interphalangeal joint, and when nothing is felt the score is 2 points. When something is felt, the vibrating tuning fork is immediately placed at the dorsal wrist. When it is felt the same at that location the score is 0 points, when it felt stronger the score is 1 point. 3) NDF/CBO 3 is combined use of 128-Hz TF testing at the hallux and SW-MF testing (see NDF/CBO 4). 4) NDF/CBO 4 is SW-MF testing. The 10-g SW-MF is tested on the plantar surface of the hallux and centrally at the heel. The ability to sense the SW-MF correctly in six trials at both locations is defined as normal; the inability to sense the SW-MF correctly in one of six trials is defined as mildly disturbed (score 1 point), and the inability to sense the SW-MF correctly more than one time is defined as disturbed and scores 2 points (11,12).

Clinical standards
Symptom scoring: DNS score. The DNS score has been described in detail elsewhere (6,13). In short, the DNS score is a four-item validated symptom score, with high predictive value to screen for PNP in diabetes (13). Symptoms of unsteadiness in walking, neuropathic pain, paraesthesia, and numbness are elicited. The presence of one symptom is scored as 1 point; the maximum score is 4 points. A score of ≥1 is defined as positive for PNP.

Physical examination scoring: DNE score. The DNE score is a sensitive and validated hierarchical scoring system (6,14). The score contains two items concerning muscle strength, one concerning reflexes, and five concerning sensation, for a total of eight items. Each item is scored from 0 to 2 (0 is normal and 2 is severely disturbed). The maximum score is 16 points. A score of ≥3 points is defined as positive for PNP (14).

Cardiovascular autonomic function testing: HRV
Cardiovascular autonomic function was assessed by analysis of continuous blood pressure and electrocardiogram (ECG) signals. In this study we used HRV as a tool reported to be sensitive and valid for diagnosis of diabetic autonomic neuropathy. All participants were studied in the morning. All measurements took place in a quiet room with the temperature kept constant at 22°C. Blood pressure was monitored by a Finapres 2300 (Ohmeda, Inglewood, CO) and heart rate by an ECG monitor (78351T; Hewlett-Packard, Palo Alto, CA). After 30 min of supine rest, the Finapres and ECG signal were sampled at 100 Hz and stored on a personal computer during 15 min. Offline 300 s of each recording was analyzed by the CARSFAPAN program (IEC ProGamma, Groningen, the Netherlands), as described previously (15,16). After correction for artifacts and a stationarity check, discrete Fourier transformation of systolic blood pressure and RR interval length was performed. For the present study, only the ECG-
derived short-term HRV analysis was performed in accordance with the guidelines of the Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (17). The total power frequency band (TP) of HRV was defined as 0.02–0.40 Hz. The reference values of the Task Force were used: HRV-TP 3,466 ms² (SD 1,018). Normal values were defined as > mean − 2 × SD; for ln HRV-TP > 7.2 is defined as normal.

**Electrodiagnostic studies: nerve conduction studies**

Nerve conduction studies were performed with standard surface stimulation and recording techniques using an electromyograph (Nicolet Viking Ile and IV) with standard filter settings. All measurements were performed after warming the forearm and lower leg in hot water (38°C) for at least 15 min. Peak-peak amplitudes were used. Reference values from our own laboratory were used, and abnormal values were defined as >2 SD of normal mean values.

Motor nerve conduction velocity (reference values) were measured in the left median (thenar) (reference value 58.5 ± 4.6 [mean ± SD] m/s) and peroneal nerves (tibialis anterior) (57.8 ± 7.1 m/s). Sensory nerve conduction velocities and amplitudes were measured antidromically with ring electrodes placed around the middle finger (median nerve) (45.6 ± SD 3.7 m/s) and stimulation lateral to the Achilles tendon (sural nerve) (47.4 ± SD 3.6 m/s). In the absence of (sensory) amplitudes, the nerve was scored as abnormal. Otherwise, the use of relative amplitudes was omitted because of the large variability in healthy subjects and diabetic patients (18). An overall NCS score was defined as the number of these four nerves with an abnormal conduction velocity, ranging from 0 (all normal) to 4 (all abnormal).

**SAC sum score**

For this study, an overall score was composed of the DNS, DNE, SW-MF, HRV, and NCS scores, representing symptom scoring, physical examination scoring, quantitative sensory testing, cardiovascular autonomic neuropathy testing, and electrodiagnostic testing, respectively, as the five categories according the SAC (5). These five tests together formed the SAC sum score. For each abnormal test result 1 point is given; the maximum score is 5 points.

**Reproducibility**

Reproducibility of the NDF/CBO 2 (tuning fork), 3 (SW-MF and tuning fork), and 4 (SW-MF) was tested in a separate study. Inter- and intrarater agreements were assessed in 13 patients. The six men and seven women, with an age of 52.5 ± 14.3 years (mean ± SD) had a wide range of neuropathy severity. The duration of diabetes was 11.6 ± 10.0 years; 3 participants had type 1 diabetes and 10 participants had type 2 diabetes. Two experienced physicians and an endocrinologist and a physician for rehabilitation medicine, both experienced in diagnosing diabetic neuropathies, rated these patients to obtain interobserver reliability; one rater observed them for a second time after 1 week to obtain intrarater reliability.

**Statistics**

The statistical package SPSS-PC version 10 (SPSS, Chicago, IL) was used to compute the descriptive statistics, Student’s t test, ANOVA, Spearman’s correlation coefficient r and Cohen’s k. The construct validity of the ICDF and NDF-CBO scores was studied with Spearman’s correlation coefficient. The discriminative power of these scores was calculated with ANOVA and independent t test.

**RESULTS**

Table 1 shows patient characteristics. Adequate matching was confirmed by a lack of differences between the groups for mean age (P = 0.15) and sex (P = 0.77) and for the DU and DC groups for the duration (P = 0.23) and type of diabetes (P = 0.33). The mean HbA1c (A1C) of the DC group was significantly lower (P < 0.01) than that of the DU group.

**Validity**

As shown in Table 2, the correlations of the scores with the clinical standards were comparable, except for the correlation of the NDF/CBO 4 score (SW-MF), which was lower. For all scores, the correlation with HRV was weaker than those with the other clinical standards.

The discriminative power, shown in Table 2, of the ICDF scores for all three patient groups was higher (F 131–134) than that of the NDF/CBO scores (F 29–118), which contain less items. The NDF/CBO 4 score had the lowest discriminative power (F 29). The discriminative power for discrimination between groups DU-DC and DU-control was comparable for all scores.

**Predictive value**

The predictive value, also shown in Table 2, was good for all scores, with the best results being obtained for the NDF/CBO 2 score (positive predictive value 86–100, negative predictive value 66–97).

**Manageability**

**Reproducibility**. For intraobserver reliability, Cohen’s k (P value) values for NDF/CBO 2, 3, and 4 scores were between 0.62 and 0.69 (P < 0.01). For interobserver reliability, Cohen’s k (P value) for NDF/CBO 2, 3, and 4 scores were between 0.60 and 0.71 (P < 0.01). Both the inter- and intraobserver reliabilities are good for all three tests.

**Performance in clinical practice**. All scores are easy to obtain in an outpatient clinic or at the patient’s bedside. Obtaining the ICDF scores takes a few minutes; obtaining the NDF/CBO scores takes <1 min.

**CONCLUSIONS** — In this study, the diagnostic scores from the ICDF and the NDF/CBO consensus have now been validated. These scores are fast and easy to obtain in clinical practice and have high predictive values to diagnose diabetic PNP. However, we showed that the single use of the 128-Hz tuning fork (NDF/CBO
generates results on validation and with predictive value similar to those obtained in combination with SW-MF and even compared with the more extended scores of the ICDF but is, of course, easier to manage. In addition to this simple test, the DNE score can be used for evaluative purposes. As has been shown earlier, the validation, predictive value, and manageability of the DNE score are good (6,13). We therefore conclude that use of the 128-Hz tuning fork can be recommended for clinical discrimination and screening for the presence of diabetic PNP.

The primary goal of this study was a comparison of the diagnostic tests of the ICDF and the NDF/CBO guidelines. After evaluation of these scores according to the criteria of Jaeschke et al. (4), we conclude that use of the tuning fork is valid and reliable for screening purposes and manageable in clinical practice. These results are confirmed by the studies of Olaleye et al. (19) and Perkins et al. (20). Perkins et al. (20) found that the SW-MF, superficial pain testing, and tuning fork by the on-off method can be confidently used for annual screening of diabetic neuropathy and that combination of the SW-MF and tuning fork does not add value to each individual screening test, which is confirmed by our findings. Olaleye et al. (19) recommended annual screening with either SW-MF or the TF by the on-off method.

Probably the panel of the ICDF chose to advise use of the tuning fork and not the vibration perception threshold (VPT) (assessed by, for example, biothesiometry) for practical reasons (7). In a separate pilot study we studied the relation of both the 128-Hz tuning fork and VPT testing (biothesiometry) with the DNE score as a clinical standard for diabetic PNP in 73 patients (having both type 1 and 2 diabetes) with a broad spectrum of duration and severity of PNP. Spearman’s correlation coefficient between the DNE score and the tuning fork and VPT scores was 0.73 ($P < 0.001$) and 0.57 ($P < 0.001$), respectively. This confirms the strength of results with the tuning fork. Therefore, we agree with the ICDF that use of the tuning fork should be preferred over VPT in diagnosing diabetic PNP not only for reasons of manageability but also because now its validity and predictive power would be even stronger by a simple test to select our diabetes control group was needed. Because the tests that were most suitable for this purpose are the subjects of this study, either as the scores we studied (for example, monofilaments or combinations of sensory tests) or as clinical standards and none of them alone can be considered as the gold standard, we had to define our group of diabetic patients without neuropathy by other parameters: absence of history of foot ulcers or clinically overt signs of neuropathy (defined as complete lack of symptoms suggestive for PNP and a normal ATR testing). Abbott et al. (10) and McNeely et al. (9) determined that the ATR testing could be used as an independent risk factor for foot ulceration due to diabetic neuropathy. Theoretically, some patients having diabetic neuropathy may have been included in this diabetic control group. However, we do not believe that this inclusion influenced the results of our study. To the contrary, the results we present about the discriminative power would be even stronger by a stronger patient selection and inclusion of diabetic neuropathy in the diabetic control group.

For this study, an overall score combining all five diagnostic categories of the SAC was needed as an alternative gold

<table>
<thead>
<tr>
<th>Score</th>
<th>ICDF 1</th>
<th>ICDF 2</th>
<th>NDF/CBO 1</th>
<th>NDF/CBO 2</th>
<th>NDF/CBO 3</th>
<th>NDF/CBO 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation: Spearman $r$</td>
<td>DNS</td>
<td>0.67</td>
<td>0.64</td>
<td>0.64</td>
<td>0.67</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>DNE</td>
<td>0.99</td>
<td>0.99</td>
<td>0.98</td>
<td>0.87</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>HRV</td>
<td>-0.44</td>
<td>-0.46</td>
<td>-0.47</td>
<td>-0.54</td>
<td>-0.50</td>
</tr>
<tr>
<td></td>
<td>NCS</td>
<td>0.61</td>
<td>0.60</td>
<td>0.60</td>
<td>0.67</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>SAC</td>
<td>0.75</td>
<td>0.74</td>
<td>0.75</td>
<td>0.83</td>
<td>0.82</td>
</tr>
<tr>
<td>ANOVA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F$</td>
<td>131</td>
<td>134</td>
<td>114</td>
<td>118</td>
<td>97</td>
<td>29</td>
</tr>
<tr>
<td>Independent $t$ test</td>
<td>DF − DC</td>
<td>11</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>DF − control</td>
<td>18</td>
<td>17</td>
<td>15</td>
<td>30</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>DC − control</td>
<td>2.5</td>
<td>2.4</td>
<td>2.4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Predictive value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNE</td>
<td>Positive</td>
<td>64</td>
<td>63</td>
<td>64</td>
<td>86</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>97</td>
<td>100</td>
</tr>
<tr>
<td>SAC</td>
<td>Positive</td>
<td>75</td>
<td>72</td>
<td>75</td>
<td>100</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>64</td>
<td>63</td>
<td>64</td>
<td>66</td>
<td>68</td>
</tr>
<tr>
<td>Manageability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reproducibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intraobserver</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interobserver</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>30 s</td>
<td>60 s</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$P$ values: Spearman correlation, all $P < 0.01$, except NCS for NDF/CBO 4, $P < 0.05$; ANOVA, all $P < 0.001$, independent $t$ test, DF − DC and DF − control, all $P < 0.001$, DC − control for IC-DF 1 + 2 and NDF/CBO 1, $P < 0.05$, NDF/CBO 2–4, NS; reproducibility, all $P < 0.01$, except NDF/CBO 3, $P < 0.001$. 
standard for diabetic PNP. Summation of the different categories might be too simple. The weight of the individual categories in the sum score is unknown. Validation of the SAC sum score was beyond the scope of this study.

In summary, use of the 128-Hz tuning fork is a valid and reliable test for screening purposes and manageable in clinical practice. The tuning fork can be used in general practice and in diabetes clinics. Use is not restricted to physicians; it can be used by nurses and paramedics too. It enables these groups to achieve their major goal: screening of large numbers of patients in their practice. The tuning fork deserves a central role in diagnosing diabetic PNP: back to basics with the tuning fork!

References