


Diagnostic accuracy of the 5.07 monofilament test for diabetes polyneuropathy: influence of age, sex, neuropathic pain and neuropathy severity

Øystein Dunker ^{1,2,3}, Martin Uglem,^{4,5} Marie Bu Kvaløy,⁶ Sissel Løseth,^{7,8} Ina Elen Hjelland,⁹ Sara Maria Allen,³ Maria Dehli Vigeland,¹ Inge Petter Kleggetveit,³ Trond Sand,^{4,5} Kristian Bernhard Nilsen^{1,3}

To cite: Dunker Ø, Uglem M, Bu Kvaløy M, *et al.* Diagnostic accuracy of the 5.07 monofilament test for diabetes polyneuropathy: influence of age, sex, neuropathic pain and neuropathy severity. *BMJ Open Diab Res Care* 2023;**11**:e003545. doi:10.1136/bmjdr-2023-003545

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjdr-2023-003545>).

Received 22 May 2023
Accepted 13 September 2023



► <http://dx.doi.org/10.1136/bmjdr-2023-003773>



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Øystein Dunker;
oeystein.dunker@gmail.com

ABSTRACT

Introduction There is a need for simple and cheap diagnostic tools for diabetic polyneuropathy (DPN). We aimed to assess the diagnostic accuracy of the 5.07/10 g monofilament test in patients referred to polyneuropathy assessments, as well as to examine how disease severity, age, sex and neuropathic pain (NP) impact diagnostic accuracy.

Research design and methods Five Norwegian university hospitals recruited patients with diabetes aged 18–70 referred to neurological outpatient clinics for polyneuropathy assessments. The 5.07/10 g Semmes-Weinstein monofilament examination (SWME) was validated against the Toronto consensus for diagnosing diabetic neuropathies; the results were stratified by age, sex and NP. Disease severity was graded by a combined nerve conduction study (NCS) Z-score, and logistic regression was applied to assess whether disease severity was a predictor of diagnostic accuracy.

Results In total, 506 patients were included in the study. Global sensitivity was 0.60 (95% CI 0.55, 0.66), specificity 0.82 (95% CI 0.75, 0.87), positive and negative predictive values were 0.86 (95% CI 0.81, 0.90) and 0.52 (95% CI 0.46, 0.58), respectively, positive and negative likelihood ratios were 3.28 (95% CI 2.37, 4.53) and 0.49 (95% CI 0.42, 0.57), respectively. The SWME was less sensitive in females (0.43), had lower specificity in patients with NP (0.56), and performed worse in patients ≥50 years. NCS-based disease severity did not affect diagnostic accuracy (OR 1.15, 95% CI 0.95, 1.40).

Conclusions This multicenter study demonstrates poor diagnostic performance for the 5.07/10 g SWME in patients with diabetes referred to polyneuropathy assessments; it is particularly unsuited for female patients and those with NP. The diagnostic accuracy of the SWME was not influenced by NCS-based disease severity, demonstrating that it does not perform better in patients with later stages of DPN. We do not recommend the use of the 5.07/10 g monofilament in the evaluation of patients with diabetes referred to polyneuropathy assessments.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ The monofilament test is widely used, but its diagnostic accuracy for diabetic polyneuropathy (DPN) is still unclear.

WHAT THIS STUDY ADDS

⇒ We show that the monofilament test is particularly unsuitable for female patients and patients with neuropathic pain, and that disease severity does not impact the test's diagnostic accuracy.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The monofilament test should not be used to diagnose DPN, nor be used as an inclusion tool in diabetes research.

INTRODUCTION

Diabetes is on the rise, now estimated to affect over 10% of the world's adult population,¹ while another 7.5% are at high risk of developing diabetes due to impaired glucose intolerance or metabolic syndrome.² Yearly global healthcare expenditures for adult patients with diabetes are nearing US\$1 trillion, or about one-tenth of the world's total health spending.¹ Diabetic polyneuropathy (DPN) is the most common complication of diabetes; it affects as many as half of patients³ and accounts for a considerable part of diabetes morbidity, mortality and financial burden.^{4–6}

There is currently no treatment for DPN. As such, the goal of early diagnosis is to delay progression of the disease and prevent related complications (eg, foot ulcers, falls or amputations), as well as help patients manage neuropathic pain and other symptoms.^{5 7–9} The most widely accepted reference standard for diagnosing DPN is nerve conduction

studies (NCS) in combination with a clinical examination.^{10–12} However, NCS is resource intensive,^{10 13} which makes it poorly suited for both small-scale and large-scale use (eg, primary care physicians or large-scale screening of diabetes populations). Thus, there is a need for simple, but sufficiently accurate diagnostic tools for DPN.

A 5.07/10 g Semmes-Weinstein monofilament examination (SWME) is commonly recommended for patients at risk for DPN.^{8 14 15} SWME is inexpensive and easy to use and is widely applied for both screening and diagnostic purposes. Nevertheless, due to large variation in test procedures and interpretations in the literature, it is still unclear whether the diagnostic accuracy of SWME is sufficient to warrant its place in the health professional's toolkit.¹⁶

The mechanical detection threshold under the feet decreases with age,¹⁷ and may also be influenced by sex¹⁸ and neuropathic pain,¹⁹ threatening the validity of a one-size-fits-all monofilament. Furthermore, it has been hypothesized that SWME predominantly detects late-stage polyneuropathy.^{10 11 20} If true, the SWME would neither contribute to early diagnosis nor facilitate early interventions, thus reducing its clinical relevance.

Therefore, the general aim of this study was to validate the Norwegian SWME protocol¹⁵ in patients with diabetes referred to polyneuropathy assessments. Specific objectives were to provide estimates for diagnostic accuracy, to examine whether DPN severity is a predictor of SWME performance, and to quantify the difference in diagnostic accuracy between sexes, age strata and for patients with neuropathic pain. This study is part of a large multicenter study including five clinical neurophysiology departments at Norwegian university hospitals.

METHODS

Overview, approval and consent

We employed a cross-sectional design to ascertain the diagnostic accuracy of the SWME in patients with diabetes, referred to polyneuropathy assessments at five different neurological outpatient clinics in Norway. The widely used Toronto consensus on diagnostic criteria for DPN served as the reference standard.¹² We graded the severity of DPN using a combined NCS score. The study follows the Standards for Reporting Diagnostic accuracy studies (STARD) guidelines.^{21 22} It is a prospective study: data collection was planned before index tests and reference tests were performed. All subjects gave informed consent prior to inclusion.

Study sample and recruitment procedure

The sample is part of a larger Norwegian multicenter study where patients aged 18–70 referred to neurological hospital outpatient clinics for polyneuropathy assessment were asked to participate, that is, a convenience sample. When a patient was referred to the hospital by a diabetes clinic or their primary care provider, they would receive an invitation letter, questionnaire and consent form by

mail. The initial exclusion criteria were acute polyneuropathies (eg, acute inflammatory demyelinating polyradiculopathy, acute motor axonal neuropathy), nerve entrapment without polyneuropathy, limited capacity to give informed consent (eg, language barrier, dementia, psychiatric illness) and patients being too sick to participate (eg, bed ridden, high fever), of which the distribution can be found in a previously published study on the same material.²³ The present study only analyzed the patients with diabetes mellitus. To reflect clinical practice, patients with likely predominant small-fiber polyneuropathy were not excluded from the main analysis.

Five hospitals participated in the data collection between May 2017 and December 2022: Oslo University Hospital; Haukeland University Hospital, Bergen; Stavanger University Hospital; St Olavs Hospital, Trondheim University Hospital, Trondheim; and University Hospital of North Norway, Tromsø.

The 5.07/10 g SWME

The SWME was performed using a 5.07 monofilament (Aesthesio, DanMic Global, California, USA) that bows at roughly '10g [axial] force' when applied at a right angle (gram-force is a deprecated, non-standard unit, equaling ~0.1 N). More accurately, it is the local *stress* or pressure that leads to the nerve response, thus the 5.07 monofilament tests cutaneous mechanical sensation at ~0.066–0.095 N/mm² (in this case Ø0.475 mm=0.17 mm²),²⁴ or ~66–95 kPa. The monofilament was replaced if it had a noticeable bend or was otherwise damaged.

The SWME followed the Norwegian guidelines by the Norwegian Directorate of Health.¹⁵ The patient was supine on an examination table for the examination. First, the physician demonstrated the feel of the monofilament by applying it to the patient's hand. Then, the patient was instructed to close their eyes, and the physician applied the monofilament again and this time asked whether the patient could feel the monofilament touching the skin. Four sites were subsequently tested on the plantar aspect of each foot, with care to avoid calluses: the heads of the first, third and fifth metatarsals, and on the distal phalanx of the first toe. The sites were tested in random order with variable rhythm. If the patient did not say 'yes' at a site, the physician came back to this site once more.

Reference standard

The reference standard was the Toronto consensus on diagnosing diabetic neuropathies.¹² The definitions of minimal criteria for DPN are: *subclinical DPN* (no signs or symptoms, but abnormal NCS or validated small-fiber test); *possible DPN*, requiring either symptoms (negative or positive symptoms; eg, numbness, pain or paresthesia) or signs (symmetrical decreased sensation or decreased/absent ankle reflexes); *probable DPN*, which requires a combination of symptoms and signs (two or more of neuropathic symptoms, decreased distal sensation, and decreased/absent ankle reflexes); and *confirmed DPN*, which necessitates the presence of an abnormality of

NCS or a validated test for small-fiber neuropathy, as well as either symptoms or signs of polyneuropathy. In accordance with the Norwegian national guidelines for clinical neurophysiology,²⁵ at least *two* nerves of different roots had to be abnormal to constitute a positive NCS finding (three if abnormal medial plantar nerve).

To avoid incorporation bias, the results of the SWME were not entered into the reference standard; instead, the clinicians used a brush, pin and/or cotton swab to test for the item 'decreased sensation'. In line with the Toronto consensus, and to reduce the number of false positives in the reference standard, only patients with *confirmed* DPN were regarded as true positives. Since confirmatory tests for DPN are not always available in real-life clinical practice, probable DPN was included as true positives in a subanalysis.

Assessment procedure

History taking, clinical examination and NCS were performed as part of routine assessments for polyneuropathy. The NCS were performed in concordance with the Norwegian national guidelines.²⁵ A minimum of two sensory nerves in the feet were tested (sural nerve and medial plantar nerve), as well as two motor nerves (tibial nerve and peroneal nerve), including F-responses. If the NCS findings were not clear, one extra sensory nerve was tested (superficial peroneal nerve).

For small-fiber neuropathy, the Toronto consensus requires normal (sural) NCS and either altered intraepidermal nerve fiber density or abnormal thermal detection thresholds in the feet. Quantitative thermal testing of the lower extremities was used as the primary confirmatory test for small-fiber lesions; skin biopsy was used in the opaque cases at Oslo University Hospital, and accounts for only 5 of 45 small-fiber diagnoses. For quantitative thermal testing, only detection thresholds were used to assess abnormality.²⁶ We employed the method of limits: the baseline temperature was 32°C, ramp rate 1°C/s and thermode size 9–12 cm², as per the national guidelines²⁵ and the hospitals' own protocol.

Local reference values for both NCS and quantitative thermal testing were applied when possible. Alternatively, reference values from Powerpack (Stefan Stålberg Software, Helsingborg, Sweden),²⁷ Hafner *et al*²⁸ or the national guidelines (quantitative thermal testing) were used.

History taking and clinical examination, including the SWME, were performed by a consultant clinical neurophysiologist. The NCS were performed on the same day as the clinical examination, most often by a technician, but were in some instances done by the consultant physician. Since the tests were conducted as part of a normal polyneuropathy assessment, the physicians were not systematically blinded from the NCS results, although at some hospitals the SWME was conducted by the physician either before NCS or without prior knowledge of the NCS results.

Grading of the severity of DPN

As a measure of DPN severity, we converted NCS results into Z-scores, including the tibial and peroneal motor nerves (distal amplitude, conduction velocity, F-M min latency), the sural nerve (amplitude and conduction velocity), the peroneal superficial nerve (amplitude and conduction velocity), and the medial tibial plantar nerve (amplitude), in line with recent work on the subject.²⁹ We then averaged the Z-scores into a single, continuous score, a *compound Z-score*. A total of 247 patients were included in this subanalysis (aged 54.8±10.7 years, 63% male) from Oslo University Hospital (n=110), Stavanger University Hospital (n=60) and Trondheim University Hospital (n=77). The reference material and Z-score calculation is described in detail in online supplemental file 1.

Neuropathic pain

Patients were assessed for neuropathic pain *in the feet* by use of the The International Association for the Study of Pain's Neuropathic Pain Special Interest Group's (NeuPSIG) criteria, described in detail elsewhere.³⁰ The NeuPSIG criteria classify patients by the level of confidence that neuropathic pain is present, graded as *unlikely*, *possible*, *probable* and *definite* neuropathic pain. Patients were dichotomized as having neuropathic pain (*probable*, *definite*) or not (*unlikely*, *possible*). We included *probable* cases in the neuropathic pain group since *definite* requires confirmatory tests that are not always readily available in the clinical setting.

Sample size

Sample size calculations for the diagnostic accuracy of the SWME were not performed, as the data for the present study were already collected as part of a larger multicenter study. However, sample size calculation for the logistic regression was conducted to ascertain how much raw NCS data on included patients needed to be retrieved from hospital databases. This calculation was based on the rule of thumb of Peduzzi *et al*,³¹ that is, $n=10 \times \text{number of covariates} / \text{smallest proportion of positive or negative cases in the population studied}$. Clinical experience suggested roughly 65% positive cases. The predictor of interest was a combined NCS Z-score but adjusted for age and sex. We therefore approximated $n=10 \times 3 / 0.35=86$ patients. When this equation provides a 'low' number, the sample size should conservatively default to 100.³² Thus, we required minimum 100 patients for the logistic regression.

Missingness, imputation and quality control of raw NCS data

In total, 11.9% of NCS data points were missing from 109 of 247 patients with incomplete studies. Of the 11 measures included, eight had missing data (tibial, peroneal and sural amplitudes were complete). Of note, peroneus superficial conduction velocity had a very high amount of missing (43.8%), partly due to non-registerable sensory nerve action potentials (20.5%),

but also as a result of its infrequent use at one of the hospitals (Stavanger). Further, the total missing of sural conduction velocity and peroneus F-waves was 16.1% and 11.6%, respectively, while the remainder were missing $\leq 10\%$ of data points. The raw NCS data were cleaned and imputed; the process is described in detail in online supplemental file 1.

Analyses of diagnostic accuracy

For the main analysis, the number of sensate sites during the examination was summed up to a total ranging from 0 (worst) to 8 (best). The patients were then dichotomized by number of sensate sites: 0–6 meant a positive test for DPN, while 7–8 meant a negative test.¹⁵ If the SWME or reference test could not be performed as per the protocol, the patient was not included in the analysis. We assessed the diagnostic accuracy of the SWME by calculating sensitivity, specificity, positive/negative predictive values (PPV/NPV), positive/negative likelihood ratios (LR+/LR-) and proportion of correctly classified patients. The overall discriminative ability of the SWME was determined by analysis of the (area under the) receiver operating characteristic curve (ROC AUC).

The clinical value of the SWME was assessed mainly through predictive values and likelihood ratios. Predictive values tell the clinician how likely it is that a given test result is true, while likelihood ratios convey how the SWME affects post-test probability of disease. For likelihood ratios, the effect of the test on post-test probability was interpreted as *small, rarely important* (1–2 and 0.5–1), *small, but sometimes important* (2–5 and 0.5–0.2), *moderate*

(5–10 and 0.1–0.2) and *large, often conclusive* (>10 and <0.01).³³ The ROC AUC value was considered to be *non-discriminative* (0.5–0.6), *poor* (0.6–0.7), *acceptable* (0.7–0.8), *excellent* (0.8–0.9) or *outstanding* (>0.9).³⁴

The effect of age, sex, neuropathic pain and DPN severity on the SWME

Contingency tables were also made for subgroups: diagnostic accuracy was calculated for males, females, patients with neuropathic pain, and the age groups <50 and ≥ 50 years old. An ROC curve was generated for each subgroup, and Youden's Index (maximum combined sensitivity+specificity–1) was marked. Logistic regression was performed to assess whether the NCS-based severity of DPN could predict the outcome of the monofilament test. The dependent variable was correct classification by the monofilament test, while the independent variable was our NCS measure of DPN severity, that is, the compound Z-score of 11 NCS measures from the lower extremities. The model was adjusted for age, sex and neuropathic pain. The Hosmer-Lemeshow test for goodness of fit was performed for each model. A p value <0.05 was considered significant.

RESULTS

In total, 506 patients were included in the study, of which 66% had confirmed DPN by the reference standard and 54% had neuropathic pain. There were no adverse events from the examinations. Demographics and clinical variables are presented in table 1.

Table 1 Patient demographics and clinical variables (n=506)

	All (n=506)	Males (n=307)	Females (n=199)	<50 years (n=139)	≥ 50 years (n=367)	Neuropathic pain (n=273)
Age, years, mean (SD)	55 (11)	57 (10)	53 (12)	41 (7)	61 (6)	56 (10)
Sex, female, n (%)	199 (39)	–	–	68 (49)	129 (35)	107 (38)
Diabetes type, n (%)						
Type 1	118 (23)	62 (20)	56 (28)	70 (50)	48 (13)	45 (16)
Type 2	321 (63)	199 (65)	122 (61)	56 (40)	265 (72)	193 (71)
Do not know	67 (13)	46 (15)	21 (11)	13 (9)	54 (15)	11 (4)
HbA1c, mmol/mol, mean (SD)	49 (50)	52 (19)	48 (22)	48 (24)	51 (18)	52 (20)
Toronto DPN consensus classification, n (%) ¹²						
Normal	43 (8)	12 (4)	31 (15)	17 (12)	22 (6)	0 (0)
Possible	67 (13)	36 (12)	31 (15)	25 (18)	42 (11)	0 (0)
Probable	64 (13)	31 (10)	33 (17)	21 (15)	47 (13)	39 (14)
Confirmed	332 (66)	228 (74)	104 (53)	76 (55)	256 (70)	234 (86)
Toronto Clinical Neuropathy Score (TCNS), mean (SD) ⁵⁴	10 (5)	10 (5)	8 (5)	8 (5)	10 (5)	12 (4)
Small-fiber neuropathy, n (%)	45 (9)	21 (7)	23 (12)	11 (8)	33 (9)	39 (14)
Neuropathic pain*, n (%) ³⁰	277 (55)	169 (55)	106 (53)	69 (50)	208 (57)	–

*Probable or definite neuropathic pain by the The International Association for the Study of Pain's Neuropathic Pain Special Interest Group's (NeuPSIG) criteria.
DPN, diabetic polyneuropathy .

Table 2 Diagnostic accuracy of the 5.07/10 g monofilament test

		Toronto consensus			Point estimate (95% CI)	
		Positive	Negative	Total		
Monofilament	Positive	200	32	232	Sensitivity	0.60 (0.55, 0.66)
	Negative	132	142	274	Specificity	0.82 (0.75, 0.87)
	Total	332	174	506	PPV	0.86 (0.81, 0.90)
					NPV	0.52 (0.46, 0.58)
					LR+	3.28 (2.37, 4.53)
					LR-	0.49 (0.42, 0.57)
					Area under the ROC curve	0.74 (0.70, 0.78)
					Correctly classified proportion	0.68 (0.63, 0.72)

LR-, negative likelihood ratio; LR+, positive likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; ROC, receiver operating characteristic.

Diagnostic accuracy of SWME for DPN

The SWME had poor sensitivity and moderate specificity for diabetes polyneuropathy (table 2). With high prevalence of confirmed DPN, the PPV was adequate, while the NPV was of little use. The modest likelihood ratios mean that the test only had a small, rarely important effect on post-test probability. The overall discriminative ability of SWME was acceptable, with a total ROC AUC of 0.74 (95% CI 0.70, 0.78) (figure 1).

The diagnostic accuracy changed when stratified for sex (table 3). On average, females were younger than

the males (53±12 years old vs 57±10 years old, p<0.01). When only females were included, SWME sensitivity was reduced by 17 percentage points, while the specificity was largely unchanged. This entails that while the PPV was still at 75% in females due to the high specificity and disease prevalence, the test now had a false negative rate of ~60%, reducing the number of true positives dramatically.

The monofilament test performed relatively similarly in age groups above (inclusive) and below 50 years. Specificity, NPV and LR+ were higher in the younger group, the latter of which approached a moderate effect on post-test probability. In patients with neuropathic pain, we found lower specificity (0.56) and subsequently lower NPV (0.20) due to an increased proportion of false positives.

For patients with predominant small-fiber neuropathy (n=45), 34% were insensate to the SWME. Excluding these patients did not result in clinically meaningful changes to diagnostic accuracy (sensitivity 0.65, specificity 0.83, PPV 0.87, NPV 0.58).

Including *probable DPN* cases as true positives had a small impact on diagnostic accuracy (sensitivity 0.56, specificity 0.91, PPV 0.96, NPV 0.36, LR+ 6.17, LR- 0.48).

Polyneuropathy severity as predictor of SWME test result

Of the 247 patients with raw NCS data, 156 (63%) had polyneuropathy by the reference standard. The median compound Z-score for the NCS measures was 0.89 (range -0.76, 6.69; IQR 2.04). NCS-based disease severity (compound NCS Z-score) was not a significant predictor of how well the SWME performed (table 4). The Hosmer-Lemeshow test for goodness of fit was non-significant for both models.

DISCUSSION

In this large, multicenter study on SWME in patients with diabetes referred to a polyneuropathy assessment, we found acceptable discriminative ability, PPV and specificity, but poor sensitivity, NPV and overall effect on

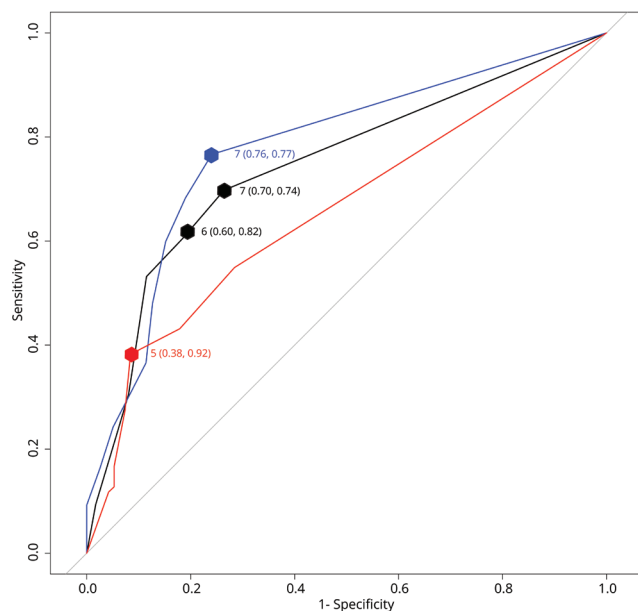


Figure 1 Area under the receiver operating characteristic curve (AUC) for the Semmes-Weinstein monofilament examination (SWME) with markers for some selected cut-offs (with their associated sensitivity and specificity in parentheses). The curve for the males is drawn in blue, females in red and the combined group in black. The preselected primary cut-off (six or less sensate sites) is marked for the combined group, as well as Youden's Index for all groups (maximum combined sensitivity+specificity-1).

Table 3 Diagnostic accuracy of the SWME by age, sex and neuropathic pain

	Point estimates (95% CI)					
	All (n=506)	Male (n=307)	Female (n=199)	<50 years (n=134)	≥50 years (n=372)	Neuropathic pain (n=273)
Sensitivity	0.60 (0.55, 0.66)	0.68 (0.62, 0.74)	0.43 (0.33, 0.53)	0.54 (0.41, 0.65)	0.62 (0.56, 0.68)	0.62 (0.55, 0.68)
Specificity	0.82 (0.75, 0.87)	0.81 (0.71, 0.89)	0.83 (0.73, 0.90)	0.89 (0.78, 0.95)	0.78 (0.69, 0.85)	0.56 (0.40, 0.72)
PPV	0.86 (0.81, 0.90)	0.91 (0.86, 0.95)	0.75 (0.62, 0.86)	0.84 (0.71, 0.94)	0.87 (0.81, 0.91)	0.90 (0.84, 0.94)
NPV	0.52 (0.46, 0.58)	0.47 (0.38, 0.56)	0.54 (0.45, 0.63)	0.63 (0.52, 0.73)	0.47 (0.40, 0.55)	0.20 (0.13, 0.28)
LR+	3.28 (2.37, 4.53)	3.60 (2.26, 5.72)	2.52 (1.49, 4.27)	4.82 (2.32, 10.01)	2.84 (1.97, 4.09)	1.42 (0.98, 2.06)
LR-	0.49 (0.42, 0.57)	0.39 (0.31, 0.49)	0.69 (0.56, 0.84)	0.52 (0.40, 0.68)	0.48 (0.40, 0.58)	0.67 (0.49, 0.93)
Area under the ROC curve	0.74 (0.70, 0.78)	0.78 (0.72, 0.83)	0.66 (0.59, 0.72)	0.72 (0.65, 0.79)	0.74 (0.70, 0.80)	0.72 (0.65, 0.79)
Correctly classified proportion	0.68 (0.63, 0.72)	0.72 (0.66, 0.77)	0.61 (0.53, 0.68)	0.70 (0.62, 0.78)	0.67 (0.62, 0.72)	0.61 (0.55, 0.67)

LR-, negative likelihood ratio; LR+, positive likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; ROC, receiver operating characteristic; SWME, Semmes-Weinstein monofilament examination.

post-test probability of disease. We showed that the 5.07 monofilament test is particularly unsuitable for female patients and patients with neuropathic pain, and that NCS-based DPN severity did not impact test performance.

The main results are in concordance with previous findings, which tend toward low sensitivity and moderate to high specificity for the SWME.¹⁶ As a simple test, the 5.07 monofilament is relatively specific, but high specificity is of little use by itself. Due to low sensitivity, almost half of patients with DPN are overlooked, diminishing the clinical value of a negative result. A possible explanation for the poor sensitivity could be our use of a sensitive reference standard. It is unlikely that a simple test for foot sensibility is as sensitive for DPN as more detailed clinical examinations, with or without NCS. Still, even simple tests should be validated against rigorous reference standards to gauge their true accuracy.

For the clinician, the test's predictive values are more important when assessing a patient than the test's credentials against a reference standard. Although there is currently no approved treatment for DPN, regular foot examinations, exercise, improved glycemic control and proper footwear are recommended management strategies.^{7 35} With early diagnosis and proper care, one can likely reduce the risk of serious sequelae from, for example, ulcers or falls. It is also important to note that although the SWME underperforms in assessing DPN, it may still be useful in predicting foot ulcers.³⁶ In addition, receiving a true DPN diagnosis empowers the patient to take an active part in the disease management. In this scenario, a moderate PPV and high NPV would be ideal.³⁷ A few more false positives would make a good trade against a reduction in false negatives, the latter of which entails that preventive action is delayed and the disease may progress further than it otherwise would have. In this light, the SWME adds little value for the clinician, since interpreting a negative result can be likened to pure chance. Therefore, we echo previous sentiments that the sole use of SWME is not sufficient for DPN assessment,^{16 35 38} neither for clinical nor research purposes. We would also question whether the diagnostic accuracy of the 5.07 monofilament supports its inclusion in larger diagnostic batteries.

The 5.07 monofilament did not perform well in females. In fact, our main result of 60% sensitivity is likely somewhat inflated due to the fact that our sample contains 50% more males than females. When stratified by sex, the proportion of correctly classified female patients is significantly reduced, with a false negative rate of well more than half. Few have investigated the effect of sex on SWME: Pambianco *et al*³⁹ did not find any difference between sexes with the 5.07 monofilament. This contrasting finding could perhaps be attributed to differences in protocol, as they only tested the dorsum of the big toe. Two studies (on the same dataset) from the German research network on neuropathic pain did not identify a significant difference on the group level between sexes for mechanical detection threshold in the

Table 4 Multivariate logistic regression: association between NCS-based disease severity and the diagnostic accuracy of the SWME

Variable	Model 1			Model 2		
	OR	95% CI	P value	OR	95% CI	P value
Compound NCS Z-score	1.20	0.99, 1.45	0.06	1.20	0.95, 1.54	0.14
Age (years)				1.01	0.98, 1.05	0.34
Sex (female)				0.80	0.41, 1.54	0.50
Neuropathic pain				0.55	0.27, 1.06	0.08

NCS, nerve conduction study; SWME, Semmes-Weinstein monofilament examination.

feet with a set of modified monofilaments.^{17 18} However, they did find higher variability and an age effect in males; since valid reference limits must account for variability, their findings suggest that different monofilaments should be used for males and females. Although there was an age difference between sexes in our study, it is unlikely to be large enough to have meaningfully impacted the results.¹⁸ This is supported by our subgroup analysis which showed little difference in overall discriminative ability between those above and below 50 years of age. If there is a true difference in mechanical detection sensitivity between the sexes, a possible solution may be to use a monofilament that exerts less pressure in female patients. Indeed, monofilaments labeled as 4.31 (~2 g) or 4.17 (~1 g) have been proposed as more sensitive tests,^{40–42} and are in fact much closer to the expected detection threshold in healthy adults.^{43–45} Comparatively, the 5.07 monofilament represents a sensory threshold tens of times greater than the expected mean in healthy individuals.^{24 44} As a result, the 5.07 monofilament lends itself to being more sensitive in patients with poorer cutaneous sensation, either as a result of advanced neuropathy or, possibly, being male. For the SWME to be worthwhile, it needs to be simple, accessible and quick to administer, which precludes the use of the 20-set monofilament kit to determine exact thresholds. Future studies may seek to properly assess the diagnostic properties of monofilaments that buckle under less force, since they are likely to increase the sensitivity in female patients.

While the most recent review of the literature did not recommend the 5.07 SWME for screening purposes,¹⁶ there has been an expectation that it would perform better in patients with more pronounced thick-fiber lesions, that is, in later stages of DPN. With the natural progression of DPN in the feet, it follows that the diagnostic accuracy of the SWME should improve in the later stages of disease.^{10 20} We found no such association between NCS-based DPN severity and the diagnostic accuracy of the SWME. Since our sample was drawn from recently referred patients, we cannot rule out that a sample with a higher incidence of severe nerve lesions could yield a different result. However, the effect is unlikely to be marked—our dataset is well populated over quite a range of DPN severity. In addition, the monofilament test is most appropriately applied early, for example, for

screening or in early diagnostic settings, when patients have similar or even less severe nerve lesions than in the present study, thus rendering the point somewhat moot.

Lastly, the SWME performed worse in patients with neuropathic pain. Although relatively few in absolute numbers, the proportion of false positives was higher than in the main analysis. This means that patients with neuropathic pain may be less sensitive to static cutaneous pressure from a monofilament at the site of pain, yet this is not directly reflected in the subanalysis' sensitivity score. One possible explanation could be that the test is affected by the presence of dysesthesia, which is common in patients with neuropathic pain.⁴⁶ Another explanation is that the findings are exaggerated, and more a result of how we dichotomize DPN (the 'false positives' have 'probable DPN'). Since many patients seek help when symptoms emerge, the clinician should be aware that symptoms of neuropathic pain may impact the SWME and interpret the results accordingly until more is known on the subject.

The present study is well powered with several strengths. It is a large multicenter study, including five Norwegian university hospitals. We employed a rigorous reference standard for diagnosing DPN, which should reduce the number of false classifications greatly. Our use of a compound Z-score for assessing DPN severity gives new insight into how the diagnostic accuracy of SWME is (un)affected by changes in nerve function.

Some limitations should also be mentioned. Our measure of DPN disease severity was based solely on NCS. This does not necessarily reflect clinical severity, which in our study was fairly low (cf Toronto Clinical Neuropathy Score). However, a compound NCS Z-score measures overall nerve function and is a good indicator of DPN progression in the myelinated nerves that respond to monofilament pressure. We did not systematically blind the consultant neurophysiologist for the NCS results, as described in detail in the Methods section. However, if some NCS results were known before the SWME, this could only have improved our results, as NCS is the far better test for large-fiber lesions. Thus, our conclusion of poor diagnostic accuracy would not change, but would be, if anything, slightly understated. Furthermore, the Norwegian SWME protocol should perhaps have been validated as a screening tool in primary care, and not as

a diagnostic tool in tertiary care. Still, if the 5.07 monofilament test cannot distinguish between patients with and without DPN in a high-prevalence setting, in the hands of a trained neurophysiologist, it is unlikely to perform better in primary care. In addition, the Norwegian SWME protocol involves testing mechanical detection thresholds at four sites on the plantar aspect of each foot. There is currently no agreed upon standard for how many sites to test, but testing a different number of sites is unlikely to meaningfully change the diagnostic accuracy.^{16–47} Lastly, we cannot be certain that the monofilaments used in this study are in fact ‘10g’, as we did not validate them on-site and throughout the testing period. On one hand, it is a known issue that the buckling force of the monofilaments may vary quite a bit from the manufacturer, and may also change over time, for example, as a response to repeated use and changes in temperature and humidity.^{48–51} On the other hand, assurances of correct buckling force would not meaningfully affect the external validity of the study, as buckling force variability is likely to be an issue everywhere.

In conclusion, the SWME with a 5.07 monofilament showed acceptable overall discriminative ability by ROC AUC, with low sensitivity and acceptable specificity in patients referred to a polyneuropathy assessment, when compared with the Toronto consensus for diagnosing DPN. The clinical value of the SWME for this population is limited: the high rate of false negatives with almost half of DPN cases missed, and the subsequent low NPV renders negative results meaningless. Overall diagnostic accuracy was similar when stratifying by age, but the SWME performed worse in patients with neuropathic pain, and in females. Future studies on SWME should properly assess whether monofilaments that exert less pressure may be more appropriate for female patients; a change in course could also be warranted as modern point-of-care devices become more affordable.^{52–53} The SWME did not perform better in patients with more pronounced thick-fiber nerve lesions. We do not recommend that the 5.07/10 g monofilament is used as a stand-alone test, nor as part of a test battery in the evaluation of patients with diabetes referred to a polyneuropathy assessment.

Author affiliations

¹Department of Research and Innovation, Oslo University Hospital, Oslo, Norway

²Faculty of Medicine, Institute of Clinical Medicine, University of Oslo, Oslo, Norway

³Department of Neurology and Clinical Neurophysiology, Oslo University Hospital, Oslo, Norway

⁴Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology Faculty of Medicine and Health Sciences, Trondheim, Norway

⁵Department of Neurology and Clinical Neurophysiology, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway

⁶Department of Neurology, Section of Clinical Neurophysiology, Stavanger University Hospital, Stavanger, Norway

⁷Department of Neurology and Clinical Neurophysiology, University Hospital of North Norway, Tromsø, Norway

⁸Department of Clinical Medicine, The Arctic University of Norway, Tromsø, Norway

⁹Department of Clinical Neurophysiology, Haukeland University Hospital, Bergen, Norway

Acknowledgements Section of Clinical Neurophysiology, Oslo University Hospital, Norway: CN technicians performing NCS and QST—Elena Petriu, Suzanna Zlateva, Ragnhild Friberg, Thomas Warvik. Physicians involved in data collection and reference test grading: Line Ulvin, Stian Hoven, Daniel Gregor Schultze, Eva Elisabeth Dornhish, Elisabeth Navjord. Department of Clinical Neurophysiology, Haukeland University Hospital, Bergen, Norway: Physicians involved in data collection and reference test grading—Anette Engelsen Eian, Sindre Rike Eng. Study coordinator and technicians: Svanhild Tøkje, Birthe Steen. St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway: Hege Michelsen, CN technician, study coordinator/study nurse. Physicians involved in data collection and reference test grading: Morten Engstrøm, Ralf Peter Michler, Arnstein Grøtting, Vibeke Arntsen, Petter Moe Omland. Additional technicians performing NCS and QST: Anne Grete Eggen, Odd Sigurd Refsnæs, Bjarte By Løfblad, Anja Skålvoll, Marthe Austvik, Lene Linn Rathe. Section of Clinical Neurophysiology, University Hospital of North Norway: Charlotte Mikkelsen (technician performing NCS and QST), Caroline Olsborg (physician involved in data collection and reference test grading). Section of Clinical Neurophysiology, Stavanger University Hospital, Norway: CN technicians taking blood samples and helping in data collection—Silje Obrestad-Aase, Irene Nessa Sæther. Physicians involved in data collection and reference test grading: Sigurbjörg Stefansdóttir, Anita Herigstad, Kira de Klerk, Nora Hognestad Haaland, Karoline Lode-Kolz, Cecilie Osman Jacobsen.

Contributors ØD, MU, TS and KBN were involved with conception and design of the study and the analysis and interpretation of the results. MU, MBK, SL, IEH, SMA, MDV, IPK, TS and KBN contributed to acquisition of the data. ØD wrote the first draft of the manuscript, and all authors edited, reviewed, and approved the final version of the manuscript. ØD is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Funding The project is funded by the Norwegian Research Council (grant number: 275476), with no involvement.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by the Regional Committees for Medical Research Ethics, South-East Norway (ref number: 2017/1593). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Qualified researchers may request access to an anonymized version of the original dataset, as well as the analysis-ready dataset for logistic regression, for up to 2 years after publication.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

ORCID iD

Øystein Dunker <http://orcid.org/0000-0003-1280-7656>

REFERENCES

- Sun H, Saeedi P, Karuranga S, *et al*. IDF diabetes atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Res Clin Pract* 2022;183:109119.
- Saeedi P, Petersohn I, Salpea P, *et al*. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the International diabetes Federation Diabetes Atlas. *Diabetes Res Clin Pract* 2019;157:107843.

- 3 Hanewinkel R, van Oijen M, Ikram MA, *et al.* The epidemiology and risk factors of chronic polyneuropathy. *Eur J Epidemiol* 2016;31:5–20.
- 4 Gordoïs A, Scuffham P, Shearer A, *et al.* The health care costs of diabetic peripheral neuropathy in the US. *Diabetes Care* 2003;26:1790–5.
- 5 Hicks CW, Selvin E. Epidemiology of peripheral neuropathy and lower extremity disease in diabetes. *Curr Diab Rep* 2019;19:86.
- 6 Hicks CW, Wang D, Matsushita K, *et al.* Peripheral neuropathy and all-cause and cardiovascular mortality in U.S. adults: a prospective cohort study. *Ann Intern Med* 2021;174:167–74.
- 7 Streckmann F, Balke M, Cavaletti G, *et al.* Exercise and neuropathy: systematic review with meta-analysis. *Sports Med* 2022;52:1043–65.
- 8 American Diabetes Association. 11. Microvascular complications and foot care: standards of medical care in diabetes—2019. *Diabetes Care* 2019;42:S124–38.
- 9 Gwathmey KG, Pearson KT. Diagnosis and management of sensory polyneuropathy. *BMJ* 2019;365:11108.
- 10 Carmichael J, Fadavi H, Ishibashi F, *et al.* Advances in screening, early diagnosis and accurate staging of diabetic neuropathy. *Front Endocrinol (Lausanne)* 2021;12:671257.
- 11 Yu Y. Gold standard for diagnosis of DPN. *Front Endocrinol (Lausanne)* 2021;12:719356.
- 12 Tesfaye S, Boulton AJM, Dyck PJ, *et al.* Diabetic neuropathies: update on definitions, diagnostic criteria, estimation of severity, and treatments. *Diabetes Care* 2010;33:2285–93.
- 13 Callaghan BC, Burke JF, Rodgers A, *et al.* Expenditures in the elderly with peripheral neuropathy: where should we focus cost-control efforts? *Neurol Clin Pract* 2013;3:421–30.
- 14 National Institute for Health and Care Excellence (NICE). Diabetic foot problems: prevention and management 2015. 2019. Available: <https://www.nice.org.uk/guidance/ng19/chapter/Recommendations#assessing-the-risk-of-developing-a-diabetic-foot-problem> [Accessed 1 Sep 2022].
- 15 Helsedirektoratet. Undersøkelse Og Diagnostikk AV Nevropati Og Identifisering AV Pasienter MED Risiko for Fotsår. 2016. Available: <https://www.helsedirektoratet.no/retningslinjer/diabetes/diabetisk-fot-og-nevropati/undersokelse-og-diagnostikk-av-nevropati-og-identifisering-av-pasienter-med-risiko-for-fotsar> [Accessed 1 Sep 2022].
- 16 Wang F, Zhang J, Yu J, *et al.* Diagnostic accuracy of monofilament tests for detecting diabetic peripheral neuropathy: a systematic review and meta-analysis. *J Diabetes Res* 2017;2017:8787261.
- 17 Rolke R, Baron R, Maier C, *et al.* Quantitative sensory testing in the German research network on neuropathic pain (DFNS): standardized protocol and reference values. *Pain* 2006;123:231–43.
- 18 Magerl W, Krumova EK, Baron R, *et al.* Reference data for quantitative sensory testing (QST): refined stratification for age and a novel method for statistical comparison of group data. *Pain* 2010;151:598–605.
- 19 Stucky CL, Mikesell AR. Cutaneous pain in disorders affecting peripheral nerves. *Neurosci Lett* 2021;765:136233.
- 20 Burgess J, Frank B, Marshall A, *et al.* Early detection of diabetic peripheral neuropathy: a focus on small nerve fibres. *Diagnostics (Basel)* 2021;11:165.
- 21 Cohen JF, Korevaar DA, Altman DG, *et al.* STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016;6:e012799.
- 22 Cohen JF, Korevaar DA, Gatsonis CA, *et al.* STARD for abstracts: essential items for reporting diagnostic accuracy studies in Journal or conference abstracts. *BMJ* 2017;358:j3751.
- 23 Dunker Ø, Grotle M, Bu Kvaløy M, *et al.* Accuracy of neuropathic pain measurements in patients with symptoms of polyneuropathy: validation of painDETECT, self-completed Leeds assessment of neuropathic symptoms and signs, and douleur neuropathique 4. *Pain* 2023;164:991–1001.
- 24 Levin S, Pearsall G, Ruderman RJ. Von Frey's method of measuring pressure sensibility in the hand: an engineering analysis of the Weinstein-Semmes pressure aesthesiometer. *J Hand Surg Am* 1978;3:211–6.
- 25 Legeforeningen. *Retningslinjer for metoder i klinisk neurofysiologi*. Legeforeningen, 2020.
- 26 Verdugo RJ, Matamala JM, Inui K, *et al.* Review of techniques useful for the assessment of sensory small fiber neuropathies: report from an IFCN expert group. *Clin Neurophysiol* 2022;136:13–38.
- 27 Stålberg S. Powerpack™. n.d. Available: <http://stalbergsoftware.com/powerpack.aspx2022>
- 28 Hafner J, Lee G, Joester J, *et al.* Thermal quantitative sensory testing: a study of 101 control subjects. *J Clin Neurosci* 2015;22:588–91.
- 29 Dunker Ø, Nilsen KB, Olsen SE, *et al.* Which combined nerve conduction study scores are best suited for polyneuropathy in diabetic patients. *Muscle Nerve* 2022;65:171–9.
- 30 Finnerup NB, Haroutounian S, Kamerman P, *et al.* Neuropathic pain: an updated grading system for research and clinical practice. *Pain* 2016;157:1599–606.
- 31 Peduzzi P, Concato J, Kemper E, *et al.* A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;49:1373–9.
- 32 Scott Long J. Regression models for categorical and limited dependent variables. In: *Advanced quantitative techniques in the social sciences*. 1997: 7.
- 33 Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients. *JAMA* 1994;271:703–7.
- 34 Hosmer DW, Lemeshow S, Sturdivant RX. *Applied logistic regression*. John Wiley & Sons, 22 March 2013.
- 35 Pop-Busui R, Boulton AJM, Feldman EL, *et al.* Diabetic neuropathy: a position statement by the American Diabetes Association. *Diabetes Care* 2017;40:136–54.
- 36 Pham H, Armstrong DG, Harvey C, *et al.* Screening techniques to identify people at high risk for diabetic foot ulceration: a prospective multicenter trial. *Diabetes Care* 2000;23:606–11.
- 37 Treveltham R. Specificity, and predictive values: foundations, pliabilitys, and pitfalls in research and practice. *Front Public Health* 2017;5:307.
- 38 Dros J, Wewerinke A, Bindels PJ, *et al.* Accuracy of monofilament testing to diagnose peripheral neuropathy: a systematic review. *Ann Fam Med* 2009;7:555–8.
- 39 Pambianco G, Costacou T, Strotmeyer E, *et al.* The assessment of clinical distal symmetric polyneuropathy in type 1 diabetes: a comparison of Methodologies from the Pittsburgh epidemiology of diabetes complications cohort. *Diabetes Res Clin Pract* 2011;92:280–7.
- 40 Brown JJ, Pribesh SL, Baskette KG, *et al.* A comparison of screening tools for the early detection of peripheral neuropathy in adults with and without type 2 diabetes. *J Diabetes Res* 2017;2017:1467213.
- 41 Kamei N, Yamane K, Nakanishi S, *et al.* Effectiveness of Semmes-Weinstein monofilament examination for diabetic peripheral neuropathy screening. *J Diabetes Complications* 2005;19:47–53.
- 42 Olaiya MT, Hanson RL, Kavava KG, *et al.* Use of graded Semmes Weinstein Monofilament testing for ascertaining peripheral neuropathy in people with and without diabetes. *Diabetes Res Clin Pract* 2019;151:1–10.
- 43 Holewski JJ, Stess RM, Graf PM, *et al.* Aesthesiometry: quantification of cutaneous pressure sensation in diabetic peripheral neuropathy. *J Rehabil Res Dev* 1988;25:1–10.
- 44 Jeng C, Michelson J, Mizel M. Sensory thresholds of normal human feet. *Foot Ankle Int* 2000;21:501–4.
- 45 Mueller MJ, Diamond JE, Delitto A, *et al.* Insensitivity, limited joint mobility, and plantar ulcers in patients with diabetes mellitus. *Phys Ther* 1989;69:453–9.
- 46 Jensen TS, Finnerup NB. Allodynia and hyperalgesia in neuropathic pain: clinical manifestations and mechanisms. *Lancet Neurol* 2014;13:924–35.
- 47 Zhang Q, Yi N, Liu S, *et al.* Easier operation and similar power of 10 G monofilament test for screening diabetic peripheral neuropathy. *J Int Med Res* 2018;46:3278–84.
- 48 Booth J, Young MJ. Differences in the performance of commercially available 10-G Monofilaments. *Diabetes Care* 2000;23:984–8.
- 49 Tan LS. The clinical use of the 10G Monofilament and its limitations: a review. *Diabetes Res Clin Pract* 2010;90:1–7.
- 50 Werner MU, Rotbøll-Nielsen P, Ellehuus-Hilmersson C. Humidity affects the performance of von Frey Monofilaments. *Acta Anaesthesiol Scand* 2011;55:577–82.
- 51 Yong R, Karas TJ, Smith KD, *et al.* The durability of the Semmes-Weinstein 5.07 monofilament. *J Foot Ankle Surg* 2000;39:34–8.
- 52 Himeno T, Kamiya H, Nakamura J. Lumos for the long trail: strategies for clinical diagnosis and severity staging for diabetic polyneuropathy and future directions. *J Diabetes Investig* 2020;11:5–16.
- 53 Kamiya H, Shibata Y, Himeno T, *et al.* Point-of-care nerve conduction device predicts the severity of diabetic polyneuropathy: a quantitative, but easy-to-use, prediction model. *J Diabetes Investig* 2021;12:583–91.
- 54 Perkins BA, Olaleye D, Zinman B, *et al.* Simple screening tests for peripheral neuropathy in the diabetes clinic. *Diabetes Care* 2001;24:250–6.

Supplemental files 1

Data-cleaning and imputation

First, two researchers (ØD and KBN) went through all the raw NCS data to identify [technical] errors in the material or measures that could not necessarily be trusted, e.g. F-waves below 30m/s in the lower limbs, motor amplitudes below 0.5 mV or registered F-waves/conduction velocities dependent on the latter. Such values had their markers in the EMG software double-checked and re-done if necessary, and patient journals were examined when required. Erroneous entries were removed and marked for imputation by the most appropriate imputation step of the two described below.

Second was the imputation process. Regular (multiple) imputation by linear regression or predictive mean matching was deemed invalid: although there was no overall discernable pattern of missingness, much of the missing data was likely pathological and therefore created a type of censored dataset (e.g. missing conduction velocity or F-waves for a nerve with non-registerable amplitudes). This meant that linear regression would underestimate the missing values and impute within a very narrow range at the model's tail, and predictive mean matching would lack valid donors. Most importantly, we had information regarding how the missing data was likely to look, which needed to be part of the imputation model for the results to make sense. Therefore, we imputed the missing data in three steps:

1. Non-registerable amplitudes were imputed with random, normally distributed values between 0 and 0.1, to avoid stacking of maximum Z-scores.
2. Missing conduction velocities and F-waves due to non-registerable amplitudes (the vast majority of missing data) were imputed as follows: For each patient, the limit of abnormality (defined as $\pm 2SD$) was calculated based on our reference material. For conduction velocities, the imputed Z-score was constrained between $Z = 2$ (positive sign is abnormal) and the Z-score of a conduction velocity 30% below the expected value (Z_{Limit})¹. For F-waves the constraints were between $Z = 2$ and an increase in latency of 20%¹. The final Z-score imputed was randomly calculated within these parameters, and then multiplied with the normal distribution, to better reflect the expected values in a large sample. For example

$$\text{Imputed } Z_{CV} = 2 + (Z_{limit} - 2) * (1 - \text{Random normally distributed value between 0 and 1})$$

Data still missing (<1%), i.e., due to technical errors or other random effects were then imputed by fully conditional specification (chained equations), with 10 iterations, based on predictive mean matching with a donor pool of five (SPSS v25). Twenty imputations were run and the average value was imputed.

Z-score calculations and grading of DPN severity

As a measure of DPN severity, we converted NCS results into Z-scores, and then averaged the Z-scores into a single, continuous score, a *compound Z-score*. A total of 247 patients were included in this sub-analysis (aged 54.8 ± 10.7 years, 63% male) from Oslo (n = 110), Stavanger (n = 60) and Trondheim (n = 77) university hospitals. The Z-scores were calculated as $Z = (\text{Observed NCS measure} - \text{predicted NCS measure}) / \text{root mean square error of the linear regression model from the reference material}$. Z-score signs were adjusted so that abnormal results (slow CVs, long latencies and low amplitudes) were always positive.

The reference material (n = 717) used for imputation and Z-score calculations was gathered between 2012 and 2022 from healthy subjects, and from patients referred to one of two Departments of Neurology in Mid-Norway (Ålesund or Trondheim) for non-specific symptoms without known disease (malignancy, diabetes, connective tissue disease, etc.), and found to be free of any neurological diagnosis after examination. Identical methods and protocols were used in both laboratories. Regression models were fitted to the data. The mean age was 45 years (SD 15 years, range 13-86 years), mean height was 171 cm (SD 7 cm, range 149-198 cm) and 70% were female. With regard to the sural nerve, differences in the NCS protocol (antidromic vs orthodromic measurements) did not allow for the reference material to be applied to patients from Oslo and Stavanger. For these patients, reference values for sural nerve conduction and sensory amplitudes were created by use of the Me-Ref method². The Me-Ref method was applied to historical records from patients referred to the two hospitals' clinical neurophysiology laboratories over the last ~20 years (*n valid measurements* ≈ 40,000). Regression models were fitted for each NCS measure, adjusted for age and height. To ensure good fit in the upper and lower ends of the model, separate regression models were made for ages 18-49 and 50-70.

The final compound Z-score included the tibial and peroneal motor nerves (distal amplitude, conduction velocity, F-M min latency), the sural nerve (amplitude and conduction velocity), the peroneal superficial nerve (amplitude and conduction velocity), and the medial tibial plantar nerve (amplitude).

1 Van den Bergh PYK, van Doorn PA, Hadden RDM, et al. European Academy of Neurology/Peripheral Nerve Society guideline on diagnosis and treatment of chronic inflammatory demyelinating polyradiculoneuropathy: Report of a joint Task Force—Second revision. *Journal of the Peripheral Nervous System* 2021;26:242-268.

2 Nandedkar SD, Stålberg EV, Barkhaus PE. MeRef: Multivariable extrapolated reference values in motor nerve conduction studies. *Muscle & Nerve* 2021;63(5):737-44. doi: <https://doi.org/10.1002/mus.27195>