Pain Physician 2013; 16:569-580 • ISSN 1533-3159

Prospective Evaluation

Assessment of Neuropathic Pain in Patients with Cancer: The Interobserver Reliability. An Observational Study in Daily Practice

Hans Timmerman, MSc[1], Irene Heemstra, MD[2], Annelies Schalkwijk, MSc[3], Constans Verhagen, MD, PhD[4], Kris Vissers, MD, PhD[5], and Yvonne Engels, PhD[6]

From: [1]Department of Anesthesiology, Pain and Palliative Medicine. Radboud University Nijmegen Medical centre, Nijmegen, the Netherlands; [2]Department of Primary and Community Care: Centre for Family Medicine, Geriatric Care and Public Health, Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands; [3]Department of Anesthesiology, Pain and Palliative Medicine, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; [4]Department of Anesthesiology, Pain and Palliative Medicine, Department of Medical Oncology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; [5]Department of Anesthesiology, Pain and Palliative Medicine, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; [6]Department of Anesthesiology, Pain and Palliative Medicine, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.

Address Correspondence:
Hans Timmerman, MSc
Radboud University Nijmegen Medical Centre
Dept. of Anesthesiology, Pain and Palliative Medicine
Huispost 717
PO Box 9101
Nijmegan, Netherlands
E-mail:
h.timmerman@anes.umcn.nl

Disclaimer: There was no external funding in the preparation of this manuscript.
Conflict of interest: Each author certifies that he or she, or a member of his or her immediate family, has no commercial association (i.e., consultancies, stock ownership, equity interest, patent/licensing arrangements, etc.) that might pose a conflict of interest in connection with the submitted manuscript.

Manuscript received: 06-14-2013
Accepted for publication: 08-13-2013

Free full manuscript:
www.painphysicianjournal.com

Background: Neuropathic pain (NeP) is a burdensome problem in all stages of cancer. Although clinical judgment is accepted as a surrogate for an objective gold standard in diagnosing NeP, no publications were found about its reliability.

Objectives: Therefore, levels of agreement on the clinical examination of NeP were estimated by calculating kappa-value (K) and percentage of pair wise agreement (PA) to determine the interobserver reliability of diagnosing NeP.

Setting: The outpatient clinic of medical oncology of the Radboud University Nijmegen Medical Centre.

Methods: Patients with cancer with potential NeP complaints were recruited from the outpatient clinic of medical oncology. Physicians were recruited from the department of pain and palliative medicine. Physicians and patients were recruited for participation in an observational study in daily practice. Each patient (N = 34) was examined by 2 specialists via independent clinical assessment. All consultations were video recorded. After each assessment, physicians were asked to indicate the most adequate characterization of the pain: pure NeP, pure nociceptive pain (NoP), mixed pain (MiP), or no pain (NP).

Results: Kappa (K) for the diagnosis of the most adequate pain characterization was 0.50, PA 64.7%. For diagnosing pure NeP k was 0.78 (PA 91.2%), for the NeP component (NeP + MiP) and NoP component (NoP + MiP), it was respectively 0.52 (PA 76.5%) and 0.61 (PA 82.4%). For the diagnosis on the basis of the grading system between physicians, K was 0.34 (PA 52.9%). The intrarater reliability for the diagnosis of an NeP component on the basis of clinical assessment and the NeP component on the basis of the grading system, for pain specialists K was 0.69 (PA 85.3%) and for palliative care specialists K was 0.61 (PA 79.4%).

Limitations: The values of K and the PA for the existence of an NeP component are not satisfying and the clinical agreement between physicians around findings from physical examination should encourage a better standardization of the clinical assessment and classification of pain in patients with cancer in respect with the identification of NeP.

Conclusions: A substantial level of agreement was found for the diagnosis of pure NeP and a moderate level of agreement for the diagnosis of the NeP component was found, both with a PA ≥ 70%. There was only a fair agreement between the physicians regarding the grading system. However, there was a substantial level of (interrater) agreement for the diagnosis of an NeP component and the outcome of the grading system. The findings in this study also suggest that a better standardization of the clinical assessment and classification of pain in patients with cancer with respect to the identification of neuropathic pain is necessary.

Key words: Neuropathic pain, diagnosis, interobserver reliability, agreement, cancer observational study, pain, clinical assessment, diagnostic test

Pain Physician 2013; 16:569-580

Pain is a burdensome symptom in all stages of cancer. Van den Beuken et al (1) found a prevalence of 55% in patients with cancer in the Netherlands. Of those, 44% suffered from moderate to severe pain (1). As described in a review, 64% of the patients with metastatic, advanced, or terminal stages of cancer had pain, 59% of patients who were on anticancer treatment and 33% of patients who had been cured from cancer still suffered from pain (1,2). In patients with cancer who were on opioid treatment by a pain specialist for their pain, almost 40% had neuropathic pain (NeP) alone or in combination with nociceptive or visceral pain (3). In several other studies, the prevalence of NeP in patients with cancer varied between 17% and 36% (4-7). This large variability in prevalence between studies can be explained by differences in populations, differences in diagnostic methodologies, and differences in definitions (8).

The International Association for the Study of Pain (IASP) defines NeP as "pain caused by a lesion or disease of the somatosensory nervous system"(9). The question arises when (part of) the pain in patients with cancer can be diagnosed as NeP. Despite the attempts to specify the entity of NeP, still no gold standard for the diagnosis of NeP exists (10). NeP is experienced by the patient and despite the characteristic signs and symptom complex that may be recognized by experienced doctors, it is still difficult to measure objectively. Several screening tools, like the DN4, LANSS, NPQ, and PainDETECT have been developed to indicate the possible existence of NeP (11-15). Yet, screening tools are no substitute for history taking and physical examination, and they are not intended to be a diagnostic method (12). Therefore, clinical judgment is the only recommended method to diagnose NeP (10,16). When standardized diagnostic criteria are lacking, the reliability of diagnostic procedures is usually demonstrated by acceptable levels of agreement among physicians (17-19). Interobserver reliability is an important measure to assess the agreement of categorical variables such as diagnosis or the interpretation of findings in physical examination (20). Cohen's kappa is a for-chance corrected statistical outcome for interobserver reliability (21). We used Cohen's kappa and percentage of pair wise agreement to investigate the interobserver reliability and agreement of the diagnosis of NeP in patients with cancer.

Methods

Patients

Patients were recruited from the outpatient clinic of the department of medical oncology of the Radboud University Nijmegen Medical Centre (RUNMC). Between September and November 2010, all patients who visited the outpatient clinic were screened for pain for another larger study. As part of a larger set of questionnaires, they were also asked to complete the 7-item DN4 questionnaire (13) about the quality of their pain. Inclusion criteria for enrollment in the kappa-study were (1) age $\geq$ 18 years; (2) diagnosed with cancer (regardless of the type and stage of cancer) or being cured from cancer; (3) at least 2 positive answers on the 7-item DN4 questionnaire in order to enrich the chance of including patients suffering from NeP in the research population.

Exclusion criteria were (1) no consent to be contacted for further research; (2) no permission for video recording of the consultations. Eligible patients were phoned by the researcher (IH). Subsequently, the patients received information by mail. After verbal and written informed consent patient-volunteers were included in the study. They did not receive any benefit from the study; only costs for transportation were reimbursed. This study was approved by the local ethics committee: the Committee on Research Involving Human Subjects region Arnhem-Nijmegen, Nijmegen, the Netherlands.

Patients were examined by experienced pain specialists (N = 4) and palliative care specialists (N = 2), recruited from the department of anesthesiology, pain, and palliative medicine of RUNMC. All 4 participating pain specialists, 2 men and 2 women, median age of 40 (range 32 – 47), had a background as an anesthesiologist. The 2 palliative care specialists, both male, were 58 and 63. One was a medical oncologist and the other an elderly care physician. Years of experience in their actual specialization (pain or palliative care) was 10 years for the pain specialists (range 1 – 18 years) and 13 years for the palliative care specialists (7 and 18 years). All physicians worked full time, but, as a mean, they worked 19 hours per week (10 – 26 hours) in this specific field.

Test Methods

All physicians completed a questionnaire recording their age, gender, professional background, specialty, and number of weekly hours working as a pain specialist or as a palliative care specialist. They

were also asked to provide a working definition of NeP, including symptoms and findings at physical examination they considered decisive for NeP. As a part of the preparation of the study, an inquiry was made among the physicians regarding the tools they wanted to use for the physical examination. There was no prearranged set of tools available in the examination rooms, only those recommended by one or more of the participating physicians: pieces of cotton wool, cotton buds, a tuning fork, and a reflex hammer. All physicians had access to the same set of tools. They were allowed to use the Electronic Patient Record (EPR), and instructed to diagnose NeP in the way they were used to in their daily practice.

Before the consultation, each patient completed a set of questionnaires, consisting of repetition of the 7-item DN4 questionnaire (13), the Brief Pain Inventory-Short Form (BPI-SF) (22), and a question about duration and course of their pain over time. Subsequently the patients were randomly assigned to be seen first by the pain specialist or the palliative care specialist and underwent a second assessment by the other specialist after 30 minutes. The physicians were not informed about the selection procedure of the participating patient-volunteers, or about the outcome of the DN4 and BPI-SF. Each physician had 20 minutes for clinical assessment of the patient (history taking and physical examination). However, the physician was allowed to take more time when necessary. After the consultation, the physician had 10 minutes to complete a research form with a tick box for the diagnosis: "NeP," "nociceptive pain (NoP)," or "mixed pain (MiP)" which was categorized as NeP together with NoP or no pain (NP). If there was more than one pain location, physicians were instructed to focus on the location of the worst pain. During the assessments, physicians were blinded to the results of their colleague and patients were instructed not to mention the findings of the other physician. In each session, 4 patients were seen in a row by each physician.

Each assessment was videotaped and evaluated by 2 researchers (IH and AS). Regarding history taking, items of evaluation were words mentioned to characterize the pain, including items of the 7-item DN4 questionnaire, and whether a Numeric Rating Scale (NRS) was mentioned (yes/no) for scoring intensity of pain. Regarding the physical examination, items of evaluation where performing a physical examination (yes/ no), comparison of affected and healthy body parts (yes/ no), and which tools were used.

Statistical Methods

Because there are no previous data regarding this research topic, it was not possible to perform a reliable power calculation. However, NeP prevalence in patients with cancer is 31% – 36% (4-7). To artificially create a higher probability of patients suffering from NeP, we included only patients who scored 2 or more items on the 7-item DN4 questionnaire during the previous screening study. We assumed NeP prevalence in this specific study group to be 0.5 during the actual study. With an assumed kappa of 0.7, a study power of 80%, and an alpha of 0.05, we estimated that 30 patients were needed. To be able to focus on agreement whether or not an NeP component exists in a patient, kappa's aimed at this specific part were determined. Patients with NeP or with MiP were rated together as having an NeP component. Patients with NoP or with MiP were also rated together as NoP component present. The physicians were, afterwards, asked to rate Treede's Grading System (23) for each patient they had seen. The outcomes "probable" and "definite" were regarded as an NeP component was present. Unlikely and possible were rated as no NeP component was present.

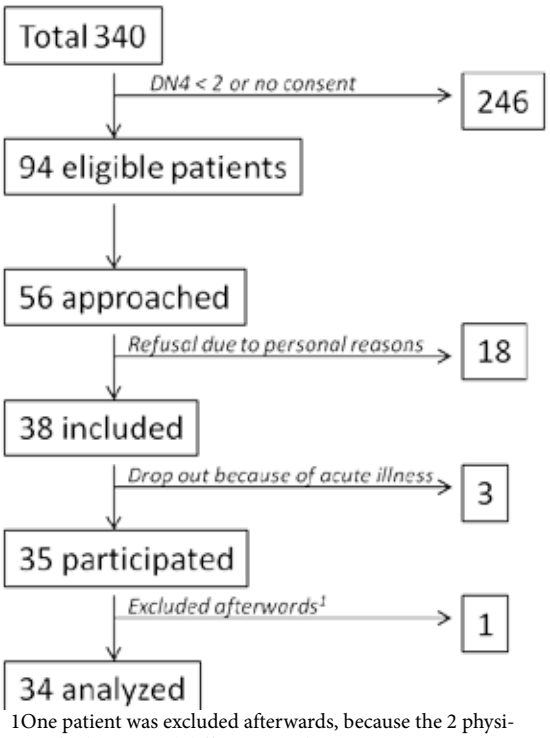
To assess interobserver reliability and agreement of the diagnosis of NeP in patients with cancer, we calculated pair-wise Cohen kappa-values (K), the prevalence index (P*i*), and pair-wise percentages of agreement (PA). K gives the proportion of agreement after chance agreement is removed (21). The K-value can vary between -1.0 and 1.0 though it usually falls between 0 and 1 (20). Landis and Koch (24) categorized values of kappa as: none beyond chance (K = 0.00), slight (K = 0.01 – 0.20), fair (K = 0.21 – 0.40), moderate (K = 0.41 – 0.60), substantial (K = 0.61 – 0.80), almost perfect agreement (K = 0.81 – 1.00). P*i* is calculated to quantify the effect of prevalence to K. It is the absolute value of the difference between the number of agreements on positive and negative findings divided by the total number of observations (20,25). PA represents the number of exact agreements divided by the number of possible agreements (26). A $\mathrm{K} \geq 0.40$ and a PA $\geq$ 70% is considered indicative of interobserver reliability acceptable for use in clinical practice (24). Statistics were applied regarding diagnosis, outcome of the grading system (23), and the outcome of the DN4. All data were entered and analyzed in Statistical Package for the Social Sciences (SPSS version 18.0, SPSS Inc., Chicago, Illinois, USA).

Results

Patients

Between September and November 2010, 340 patients visiting the outpatient clinic of the department of medical oncology of the RUNMC completed the pain questionnaire. Of them, 94 scored 2 or more on the 7-item DN4 and gave their consent to be approached for a subsequent pain study (Fig. 1). After 56 patients were approached we stopped the inclusion in this study. Eighteen patients refused to join the study due to personal reasons (mainly because of active ongoing chemotherapy schedules). Finally, 38 patients gave their written informed consent. Due to an acute intercurrent illness at the day of the assessments, 3 patients dropped out of the study. Therefore, 35 patients participated in the kappa study. One patient was excluded afterwards, because the 2 physicians had examined different pain locations.

These 34 patients had a median age of 56 (range 36 – 76). There were 8 men (24%), of whom 2 had testis carcinoma, 4 had tumors arising from the digestive system, one had a GIST tumor, and one had a carcinoid. Of the 26 women (76%), 92% had breast cancer (N = 24), one a GIST tumor, and one an angiosarcoma. The duration of the pain in months was at mean 64 months (± SD 100; range 1 – 568 months). Worst pain during the 24 hours before the consultations was experienced as severe in 5 cases (15%), moderate in 21 cases (63%), and mild in 7 cases (7%): mean 5.24 ± SD 2.28; range 0 – 9 (NRS 0 – 10). The average pain in the last 24 hours was at mean 4.19 ± SD 2.15; range 0 – 9 (NRS 0 – 10). The outcome of the BPI-SF for the pain severity score at mean was 4.08 ± SD 2.23; range 0 – 8 (NRS 0 – 10) and for the pain interference score 3.67 ± SD 2.37; range 0 – 9 (NRS 0 – 10). On the repeated 7-item DN4 questionnaire on the day of examination, one patient didn't fill in the questionnaire, one patient scored 0 points, 8 patients scored 2 points, 11 scored 3 points, 10 scored 4 points, 2 patients scored 6 points, and one 7 points. See Table 1 for more detailed patient characteristics.



1One patient was excluded afterwards, because the 2 physicians had examined different pain locations.

Fig. 1. Flowchart for inclusion of patients.

Physicians

We asked the physicians, in an open question, to give their working definition of NeP: 2 of the pain specialists mentioned the definition suggested by Treede (23), one pain specialist mentioned the DN4-criteria, and the other physicians mentioned definitions containing the words "pain" and "the nervous system/ nerve damage." To the question "what do you think is a decisive symptom for NeP," 3 pain specialists answered that allodynia in general was the decisive symptom and one had the opinion that there was none. The palliative care specialists considered respectively a changed sensibility and an annoying pain during night the decisive symptom. When asked for the decisive finding for NeP at physical examination, again allodynia was mostly mentioned by the pain specialists, while the palliative care specialists mentioned changed sensibility and hyperpathy (Table 2).

Test Results

The K and PA between paired physicians for the characterization of pain (NeP, MiP, NoP, or NP) was 0.50 (64.7%) ($P < 0.000$). For diagnosing NeP K was 0.78, P*i* 0.44, and PA 91.2%; for MiP it was respectively 0.53, 0.38, and 79.4%; and for NoP it was 0.31, 0.26, and 67.6%. The K for the NeP component (by summing the diagnoses of pure NeP and MiP) was 0.52 ($P = 0.002$), P*i* was 0.18, and PA was 76.5%. For the NoP component (by summing the diagnoses of pure NoP and MiP) K was 0.61, P*i* was 0.35, and PA was 82.4%.

Table 1. *Summary of the patient characteristics and outcomes of the assessments.*

Gender	Age	Primary type of cancer	Disease stage	Anti-cancer treatment	Worst pain score patient	7-item DN4 score	Physical examination		Clinical cause of pain		Pain diagnosis		Grading system	
							SPC	PS	SPC	PS	SPC	PS	SPC	PS
F	65	Breast	1	S, RT	6	1	1,2,4,8	1,7	Postoperative neuropathy	Triggerpoint pain	NeP	NoP	D	U
M	56	Rectum	2	RT	m	m	1,2,4,8	1	Postradiation pain peri-anal	Triggerpoint pain	NeP	NoP	D	U
F	57	Breast	2	S,RT	6	4	1,2,4,6,8	0	Arthrosis cervical spine	Cervical facet joint pain	NoP	MiP	U	P
F	59	Breast	1	S, RT, C	5	4	1,2,4,6,7,8	2,4,6,7,8	Post surgical nerve injury	Surgery and Chemotherapy	MiP	MiP	P	Po
F	52	Breast	3	S, C	3	3	1,2,4,8	1,2,3,7,8	Metastatic collapse thoracal and lumbar spine	Vertebral metastases	NoP	NoP	D	U
F	56	Breast	2	S, RT, C	5	4	7,8	1,4,6,7,8	Arthrosis	Rheumatoid arthrosis	NoP	NoP	D	U
M	68	Rectum	3	S, C	8	4	2,6,7,8	1,2,4,6,8	Oxaliplatin	Chemotherapy	NeP	NeP	P	P
F	50	Breast	1	S, RT, C	6	6	1,2,7,8	1,2,8	(1)Postoperative neuropathy after mama amputation (2) allodynia	Periost pain costa	NeP	NeP	D	P
F	53	Breast	2	S, C	4	3	0	1,4,8	Muscle - and joint-pain	(1) Rheumatoid arthritis or arthrosis (2) Tamoxifen side-effects	NoP	MiP	D	D
M	36	Testis	1	S,C	9	2	1,2,4,6,8	2,3,6,7,8	Unclear	Unknown	NeP	NeP	D	Po
M	76	Carcinoid	3	S,C	7	4	1,8	2,3,8	Postoperative neuropathy	Intercostal nerve damage with neuralgia	NeP	NeP	P	D
F	59	Breast	3	S, RT, C	9	4	4,7,8	2,3,7,8	(1) Degenerative disorders (2) multiple sclerosis	Vertebral metastasis	MiP	MiP	P	U
F	52	Breast	2	S, RT, C	5	3	1,2,4,8	2,6,8	(1) Arthrosis (2)use of Tamoxifen	Degenerative disorder back	NoP	MiP	U	U
F	58	Breast	1	S,C	0	2	1,2,4,6,7	2,3,6,8	No pain	Prothesis	NP	NoP	U	U
F	50	Breast	3	S, RT, C	7	2	1,2,8	7	Bone metastasis costae	Visceral bone pain	NoP	NoP	U	U
F	65	Breast	3	S, C	4	6	1,2,4,6	2,3,6,8	Myogene disorders + Fear	Chemotherapy, surgery, backpain and medication	NoP	MiP	P	D
F	60	Breast	3	S, C	9	4	1, 8	1,2,6,7	Vertebral metastasis T10/T11	Metastatic disorders with intercostal neuralgia T11-T12	MiP	MiP	D	D
F	62	Breast	3	S, RT, C	5	3	1,2,4,6,8	2,3,6,7,8	(1)Arm pain: post surgery (2) Hand+Feet pain post taxol	Scar tissue	NP	NoP	P	U
F	49	Breast	2	S, C	4	4	1,2,4,6,8	2,3,6,8	Chemotherapy	Allodynia, sensibility disorder	NeP	NeP	P	P
F	46	Breast	2	S, RT, C	7	3	1,2,4,8	1,4,6,8	Post operative, post chemotherapy, nociceptive pain by lymphedema	(1)Axillair lymphe dissection, (2) Radiotherapy, (3)Chemotherapy, (4)Migraine	MiP	MiP	D	D
F	54	Breast	2	S, C	5	3	1,2,4,6,7,8	2,7,8	Breast surgery	Post surgical scar pain	NeP	NeP	P	P
F	52	Breast	2	S, RT, C	6	2	8	2,7,8	Anti-oestrogen treatment	Reactive arthritis complaints following hormonal treatment	NoP	MiP	U	U

Table 1 (cont.). *Summary of the patient characteristics and outcomes of the assessments.*

Gender	Age	Primary type of cancer	Disease stage	Anti-cancer treatment	Worst pain score patient	7-item DN4 score	Physical examination		Clinical cause of pain		Pain diagnosis		Grading system	
							SPC	PS	SPC	PS	SPC	PS	SPC	PS
F	61	GIST	3	S, C	8	3	0	2,3,7,8	Missing	Myotendinogen lumbar pain/ sacro iliac joint pain	NoP	NoP	U	U
F	57	Breast	1	S, RT, C	3	2	8	1,4,8	(1)epicondylitis lateralis (2)Ablatio, preoperative neuropathy	Multiple surgery	MiP	MiP	D	D
F	45	Angio-sarcoma	3	C	7	3	1,2,4,6,7,8	0	Bone metastasis spine	Lumbar facet pain with pseudoradicular signs	MiP	MiP	P	U
F	70	Breast	3	S, RT, C	5	3	7,8	0	(1)Plexus brachialis leasion (2) post radiation	(1)plexopathy (2)m.raynaud	NeP	MiP	D	D
M	69	Rectum	2	S, RT, C	3	0	1,4,7	1,2,4,8	Degenerative disorders lumbosacral spine	Discopathy lumbar spine with radicular signs	MiP	MiP	D	U
M	56	GIST	4	S	6	2	1,2,4,7,8	7	Tumor localisation	Oncologic process, GIST tumor calve	NoP	NoP	U	U
F	55	Breast	1	S, RT, C	7	4	1,2,4,8	7	(1)Surgery (2)Edema	Edema	MiP	NoP	P	U
M	65	Rectum	2	S, RT, C	4	3	0	0	No pain	Postoperative scar tissue pain	NP	NoP	Po	U
F	52	Breast	3	S, RT, C	5	2	1,2,6,7	7,8	Bone metastasis mama carcinoom	Bone metastasis T5	NoP	NoP	U	U
F	70	Breast	3	S, C	1	7	1,2,4,6,8	0	Chemotherapy	Chemotherapy induced neuropathy	NeP	NeP	P	P
M	45	Testis	1	S, C	3	4	1,2,4,6,7,8	2,8	Cisplatinum	Chemotherapy induced neuropathy	NeP	NeP	P	P
F	44	Breast	1	S, C	1	3	0	0	Surgery	Incission by surgeon	NoP	NoP	U	U

Gender: M = male, F = female; Disease stage: 1 = patients who received anti-cancer treatment with curative intent ≥ 6 months ago, 2 = patients receiving anti-cancer treatment with curative intent or last treatment less than 6 months ago, 3 = patients receiving palliative anti-cancer treatment, 4 = no treatment or treatment no longer feasible; Anti-cancer treatment: Previous anti-cancer therapies S = surgery, RT = radiotherapy, C = chemotherapy, O = other; NRS worst pain patient: Worst pain during 24 hours before consultation: Numeric Rating Scale (NRS) 0 – 10, m = missing; 7-item DN4 score: 7-item DN4 score at day of assessment on all pain locations, m = missing; Pain diagnose: NoP = nociceptive pain, NeP = neuropathic pain, MiP = mixed pain (nociceptive and neuropathic pain), NP = no pain; PS: Pain Specialist; SPC: Specialist Palliative Care; Physical examination: tools or techniques used during physical examination: 0 = none, 1 = soft touch, 2 = pin prick, 3 = piece of cotton wool, 4 = tendon reflexes (reflex hammer), 6 = tuning fork, 7 = stroking over skin with hands, 8 = comparing affected and healthy body parts. P = probable. U = unlikely. D= definite. Po = Possible

Table 2. *Individual opinions from participating physicians about diagnosing neuropathic pain in general.*

	Specialists Palliative Care	Pain Specialists
Working definition for neuropathic pain	Pain or troublesome experience of the patient that can be traced back to a possible or demonstrated change in the function of the nerve or central nervous system.	Pain arising as a direct consequence of a lesion or disease affecting the somatosensory system (Treede, 2008). (2x)
	Pain related to the peripheral or central nervous system.	Pain as a consequence of nerve damage or neurological dysfunction including sensitization.
		DN4-criteria.
Decisive symptom for neuropathic pain	Changed sensibility (experienced as pain/ troublesome).	Pain at normal touch.
		None.
		Allodynia.
	Especially pain during nighttime, mostly annoying.	Allodynia (dynamic and static) and abnormal sensations.
Decisive sign at physical examination for neuropathic pain	Changed sensibility in an area of pain experience.	Allodynia. (2x)
		Allodynia static and dynamic.
	Hyperpathy.	Sensorial abnormality.

Table 3. *The kappa coefficient (*K*) and the percentage of pair-wise (PA) agreement between physicians calculated for the patients' diagnosis.*

	K-value	Approx. Sig.	Categorized value of kappa	P*i*	PA-value (%)
Diagnosis (NeP, MiP, NoP, NP)	0.50	0.000*	Moderate	---	64.7
NeP (NeP versus MiP + NoP + NP)	0.78	0.000*	Substantial	0,44	91.2
MiP (MiP versus NeP + NoP + NP)	0.53	0.001*	Moderate	0,38	79.4
NoP (NoP versus NeP + MiP + NP)	0.31	0.08	Fair	0,26	67.6
NePcomponent (NeP + MiP versus NoP + NP)	0.52	0.002*	Moderate	0.18	76.5
NoPcomponent (NoP + MiP versus NeP + NP)	0.61	0.000*	Substantial	0.35	82.4

κ - value: Kappa value; 95% CI: 95% confidence interval; Approx. Sig.: Approximate significance; *: significant, $P \leq 0{,}05$; P*i*: Prevalence index; PA-value: Pair-wise Agreement-value; NeP: neuropathic pain; MiP: mixed pain; NoP: nociceptive pain; NP: no pain; Fair: $\kappa = 0.21 - 0.40$; Moderate: $\kappa = 0.41 - 0.60$; Substantial: $\kappa = 0.61 - 0.80$.

The interobserver reliability and the pair-wise agreement between the pain specialist and the palliative care specialists regarding the grading system (unlikely, possible, probable, and definite neuropathic pain) showed a K of 0.34 and a PA of 52.9%. The comparison between the NeP component, following from the diagnosis of the physician and the outcome of the grading system (the outcomes probable and definite were regarded as an NeP component was present) gave for the pain specialists a K of 0.69, P*i* 0.26, and PA of 85.3%. For the palliative care specialists it was respectively 0.61, 0.03, and 79.4%. The comparison between the NeP component, following from the diagnosis of the physician and the outcome of the DN4 (7-items, a "yes" on ≥ 3 items is considered as having NeP) gave for the pain specialists a K of 0.24, P*i* 0.36, and PA of 66.7%. For the palliative care specialists it was respectively 0.16, 0.27, and 57.6%. The comparison between the outcome of the grading system (the outcomes probable and definite were regarded as an NeP component was present) and the outcome of the DN4 (7-items, a "yes" on ≥ 3 items is considered as having NeP) gave for the pain specialists a K of 0.34, P*i* 0.42, and PA of 72.7%. For the palliative care specialists it was respectively 0.32, 0.15, and 63.6%.

Secondly, items from history taking and physical examination were assessed by video recording and

Table 4. *The kappa coefficient (*K*) and the percentage of pair-wise agreement (PA), calculated for the NeP component, grading system, and DN4.*

	K-value	Approx. Sig.	Categorized value of kappa	P*i*	PA-value (%)
Grading PS & Grading SPC (unlikely-possible-probable-definite)	0.34	0.001*	Fair	---	52.9
PS: NePcomponent & Grading NePcomponent (NeP + MiP versus Grading probable + definite)	0.69	0.000*	Substantial	0.26	85.3
SPC: NePcomponent & Grading NePcomponent (NeP + MiP versus Grading probable + definite)	0.61	0.000*	Substantial	0.03	79.4
PS: NePcomponent & DN4 (NeP + MiP versus DN4)	0.24	0.160	Fair	0.36	66.7
SPC: NePcomponent & DN4 (NeP + MiP versus DN4)	0.16	0.475	Slight	0.27	57.6
PS: Grading NePcomponent & DN4 (Grading probable + definite versus DN4)	0.34	0.053	Fair	0.42	72.7
SPC: Grading NePcomponent & DN4 (Grading probable + definite versus DN4)	0.32	0.026*	Fair	0.15	63.6

κ-value: Kappa value; 95% CI: 95% confidence interval; Approx. Sig.: Approximate significance; *: significant, $P \leq 0,05$; Pi: Prevalence index; PA-value: Pair-wise Agreement-value; PS: Pain Specialist; SPC: Specialist Palliative Care; DN4: Douleur Neuropatique en 4 Questions (Questionnaire); NeP component: Neuropathic pain component (diagnosis NeP or MiP); Grading NeP component: only "probable" and "definite" are counted as NeP component; MiP: mixed pain; NoP: nociceptive pain; Grading: Grading system by Treede et al (23); Slight: $\kappa = 0.10 - 0.20$; Fair: $\kappa = 0.21 - 0.40$; Moderate: $\kappa = 0.41 - 0.60$; Substantial: $\kappa = 0.61 - 0.80$.

analyzed afterwards. In 27 out of 34 cases the palliative care specialists asked for a pain score and the pain specialists asked in 21 cases. Most frequently asked items of the DN4 during history taking were tingling (23 times by the palliative care specialists and 15 times by the pain specialists), numbness (18 times by both), and burning (12 versus 19 times). During physical examination, the cotton bud was most often used. The palliative care specialists used the sharp side of a cotton tip 22 times and the pain specialists 18 times. The soft side of it was used 25 times by the palliative care specialists and 10 times by the pain specialists. Of the available tools, the cotton wool was used the least: 9 times by the pain specialists while the palliative care specialists did not use it at all.

Discussion

In this real-life type of study, we found a substantial level of interobserver reliability for diagnosing pure NeP and a moderate level of interobserver reliability for the diagnosis of an NeP component, between pain specialists and specialists in palliative care, both with a $K \geq 0.40$ and a $PA \geq 70\%$. A K of ≥ .40 and a PA of ≥ 70% is indicative of interobserver reliability and acceptable for clinical use (25). The reliability of the diagnosis of NeP by a physician is an important consideration in clinical practice because it has direct treatment implications for the individual patient. We performed this kappa study to see if the diagnosis of NeP is a reliable diagnosis because an objective gold standard for this diagnosis is lacking. As an example, in validation studies of questionnaires screening for NeP 2 physicians were both examining the same patient to serve as a substitute gold standard for diagnosis (13,14). But until now no proof of this concept was given. According to the literature (20,25,27,28) we chose to use the kappa-value as well as the PA and P*i*. The level of agreement for NeP component either as a part of MiP or as pure NeP appeared moderate. Regarding MiP we found a moderate but significant level of agreement which suggests that the clinical picture is less straight forward. Probably, a combined pain syndrome is a less clear outcome, explaining the lower kappa. For pure NoP the physician pairs only had a fair, non significant level of agreement. The PA-value for NoP was below 70% and thus considered as not acceptable for clinical use. This might be due to the focus of the physicians: the instruction of the physician was to diagnose NeP in the way they were used to in their daily practice. Probably there was less attention to NoP. For NoP component the level of agreement was substantial.

Although the participating physicians used different descriptions for NeP, a high consensus existed for the decisive symptom and sign for NeP, namely allodynia or a description of allodynia. However, allodynia is not a decisive symptom for NeP, because it might also

be present in patients with nociceptive pain, especially in inflammatory conditions.

The presented results indicate that the specialists used very different diagnostic criteria for neuropathic pain. This was confirmed most notably by the working definition used by the investigators, which corresponded to the IASP definition of neuropathic pain in only one third of the investigators (Table 2). In conclusion, the majority of the participating physicians didn't know the current definition of neuropathic pain and use "personalized" inappropriate diagnostic criteria in their daily practice.

In this study, we also have used the grading system by Treede et al (23), filled in by both physicians after the clinical examination of the patient. Comparing the diagnosis of the existence of an NeP component with the outcome of the grading system per physician, we found a substantial intraobserver reliability with a PA ≥ 70%, indicating a good reliability and useful in clinical practice. However, the comparison on the outcome of the grading system (unlikely, possible, probable, or definite) between both physicians gave a fair reliability and a low PA (< 70%), indicating a poor reliability between both physicians and therefore it might be less useful in clinical practice. Moreover, the grading system will not necessarily provide the right diagnosis. In a patient suffering from MiP, the NoP part may be paramount. The physicians' diagnosis (NeP, MiP, NoP, or NP) had a moderate reliability, but also a low PA < 70%. All this indicates that it is difficult to categorize the kind of pain the patient is suffering from, as well with the physicians' diagnosis as with the grading system. It can be questioned whether the clinical judgment should be regarded as a gold standard for the diagnosis of NeP because both clinicians might be wrong in their diagnosis even with values of $\kappa > 0.5$ and a PA of 70%.

Our study measured the interobserver reliability of 2 physicians diagnosing NeP in patients with pain from cancer and taking the grading system and the DN4 into account. The focus of the study was to diagnose the kind of pain and not on which specific diagnostic tests were used in the diagnostic process. Mostly, kappa studies are used to report the reliability of specific diagnostic tests in patients or from clinical data (18,29-31). Comparing the outcome of the physicians diagnosis on the existence of an NeP component with the outcome of the 7-item DN4 we found only a fair ($\kappa < 0.40$) reliability for the pain specialists and a slight interobserver reliability for the palliative care specialists (both with a PA ≤ 70%). In the paper of Garcia de Paredes et al (7) it was described that only half of the patients with cancer suffering from NeP had a positive score on the DN4 compared with the clinicians diagnosis. They suggested investigating if a specific cut-off score for the DN4 for patients with NeP from cancer would fit better. The same was suggested in the study by Mercadente et al (32) for the LANSS, NPQ, and NPQ-SF. This study also indicates that the DN4, at this moment, is less valid and thus less useful in clinical practice for screening for NeP in patients with cancer pain.

During the pain history taking, the pain specialists asked for a pain intensity score only 21 out of 34 times and the palliative care specialists 27 times. A marginal comment should be made on this statement, as the physicians were only instructed to diagnose the type of pain. However, one expects a pain intensity score to be a standard item during a pain history taking. During the observation of the clinical examination of the patients, in 10 of 68 cases (Table 1), no clinical examination was performed, and in many cases only one sensory modality was tested. It has been recommended (16) that (a) clinical bedside (sensory) examination of a patient with suspected NeP includes testing of touch/vibration, cold, warmth, and pain sensibility (pinprick) and (b) the outcomes should be compared with the findings in the contra lateral region or in a region without pain (not performed in 21 of 68 cases).

To our knowledge, this is the first study to determine the interobserver reliability of the diagnosis of NeP in patients with cancer. Participation of patients, examination rooms that were equipped as real consultation rooms, and instructing the physicians to perform the diagnosing procedure as they usually do, all contributed to collect reliable information about the current state of daily practice in this hospital. Besides, by using a video camera that was almost invisible to the physician and patient, the consultation was not disturbed by the researchers.

While the patient number (N = 34) is sufficient for a reliable kappa study, the number of participating physicians was low and unequal: 4 participating pain specialists and 2 palliative care specialists. Another weak aspect was the fact that one of the palliative care specialists was a medical oncologist and the other an elderly care physician. Yet, both of them had palliative care as their main task for at least 7 years. However, both the pain specialists and the palliative care specialists will be more experienced than usual physicians in pain and NeP and our findings cannot be interpreted for a broader group of physicians. Further-

more, in the questionnaire, patients were asked about all sites of pain; whereas in the clinical examination, physicians were instructed to focus on the site of worst pain. The majority of patients (25) had breast cancer. This high number is an adequate representation as breast cancer is the most frequent type of cancer among women in the Netherlands (www.cijfersoverkanker.nl) and many of them suffer from chronic pain (33,34). The incidence of NeP in this study is artificially high in comparison with the normal population in the oncology outpatient clinic. Because we used a score of at least 2-points on the 7-item DN4 as an inclusion criterion for this study, the presence of NeP was more likely and thus enlarged the possibility of diagnosing NeP. For now it is not sure that in a situation of a lower incidence of NeP the kappa values will be the same. The physicians were also more triggered and focused on NeP than on NoP because we asked their working definition of NeP, the symptoms and findings at physical examination they considered decisive for NeP, and their self-efficacy in diagnosing NeP. This is probably the cause of a lower kappa-value in patients with NoP. Finally, the worst pain did not necessarily originate from the cancer or anti-cancer treatment. Patients sometimes had comorbidity causing the (worst) pain, for example rheumatoid arthritis.

Conclusions

We found a substantial level of agreement for the diagnosis of NeP and a moderate level of agreement for diagnosing an NeP component, both with a PA $\geq$ 70%. This study shows preliminary evidence that the clinical judgment of NeP in patients with cancer is reliable. Implementation of the proposed criteria for categorizing NeP as definite, probable, possible, or unlikely might be a step forward (23) to come to more diagnostic clarity for NeP. As stated by Bennett et al (8) a standardized approach is essential for clinical assessment, for appropriate treatment, and for clinical research. Despite the lack of a gold standard for diagnosing NeP, our study shows that physicians have a good agreement in the diagnosis of pure NeP. For MiP however, the level of agreement is moderate but with a high PA. Based on these findings, in MiP we suggest the opinion of a second physician to enlarge the chance of a correct diagnosis and thereby of adequate pain treatment. Especially in more complex pain syndromes, the recognition of NeP component needs attention. Since the treatment of NeP and MiP or NoP is quite different according to the international guidelines, a strict delineation and certitude about the correct diagnosis is of upmost importance and will influence the result of consequent pharmacological treatment schemes (35). Taking into account the different pain mechanisms of NeP and NoP and working mechanisms of the medications, it is important to have an adequate pain diagnosis for optimal pain treatment with the least side effects. The general value of the findings for validating physician assessment of neuropathic cancer pain in this study is limited to our centre and participating physicians in order to confirm their relevance and general interest. However, the findings in this study suggest that a better standardization of the clinical assessment and classification of pain in patients with cancer in respect to the identification of neuropathic pain is necessary. Moreover, we recommend a further study on how to improve the level of agreement in, and the validity of, the clinical diagnosis of NeP by systematically analyzing the history taking and the different (diagnostic) tools used in pain assessment and how standardizing the diagnostic process can improve the level of agreement and validity in clinical circumstances (16,23).

Acknowledgments

We thank the participating patients and physicians for their contribution to this study.

Funding

This study was co-funded by the Netherlands Organisation for Health Research and Development. No specific conflicts of interest are present for the authors.

Author Contributions

HT, IH, AS, SV, KV, and YE designed the trial protocol. KV and YE secured funding for the kappa study. AS and IH performed this study. HT, IH, and AS drafted the manuscript. YE, SV, and KV contributed to the manuscript. All authors read, discussed the results, and commented on the manuscript. At the end, all authors approved this final manuscript.

References

1. van den Beuken-van Everdingen MH, de Rijke JM, Kessels AG, Schouten HC, van Kleef M, Patijn J. High prevalence of pain in patients with cancer in a large population-based study in The Netherlands. *Pain* 2007; 132:312-320.
2. van den Beuken-van Everdingen MH, de Rijke JM, Kessels AG, Schouten HC, van Kleef M, Patijn J. Prevalence of pain in patients with cancer: a systematic review of the past 40 years. *Ann Oncol* 2007; 18:1437-1449.
3. Caraceni A, Portenoy RK. An international survey of cancer pain characteristics and syndromes. IASP Task Force on Cancer Pain. International Association for the Study of Pain. *Pain* 1999; 82:263-274.
4. Grond S, Radbruch L, Meuser T, Sabatowski R, Loick G, Lehmann KA. Assessment and treatment of neuropathic cancer pain following WHO guidelines. *Pain* 1999; 79:15-20.
5. Grond S, Zech D, Diefenbach C, Radbruch L, Lehmann KA. Assessment of cancer pain: a prospective evaluation in 2266 cancer patients referred to a pain service. *Pain* 1996; 64:107-114.
6. Manas A, Monroy JL, Ramos AA, Cano C, Lopez-Gomez V, Masramon X, Perez M. Prevalence of neuropathic pain in radiotherapy oncology units. *Int J Radiat Oncol Biol Phys* 2011; 81:511-520.
7. Garcia de Paredes ML, del Moral Gonzalez F, Martinez del Prado P, Marti Ciriquian JL, Enrech Frances S, Cobo Dols M, Esteban Gonzalez E, Ortega Granados AL, Majem Tarruella M, Cumplido Buron JD, Gasco Hernandez A, Lopez Miranda E, Ciria Santos JP, Castro Carpeno FJ. First evidence of oncologic neuropathic pain prevalence after screening 8615 cancer patients. Results of the On study. *Ann Oncol* 2011; 22:924-930.
8. Bennett MI, Rayment C, Hjermstad M, Aass N, Caraceni A, Kaasa S. Prevalence and aetiology of neuropathic pain in cancer patients: a systematic review. *Pain* 2012; 153:359-365.
9. The International Association for the Study of Pain (IASP): http://www.iasp-pain.org/Content/NavigationMenu/GeneralResourceLinks/PainDefinitions/default.htm
10. Freynhagen R, Bennett MI. Diagnosis and management of neuropathic pain. *BMJ* 2009; 339:b3002.
11. Krause SJ, Backonja MM. Development of a neuropathic pain questionnaire. *Clin J Pain* 2003; 19:306-314.
12. Bennett MI, Attal N, Backonja MM, Baron R, Bouhassira D, Freynhagen R, Scholz J, Tölle TR, Wittchen HU, Jensen ST. Using screening tools to identify neuropathic pain. *Pain* 2007; 127:199-203.
13. Bouhassira D, Attal N, Alchaar H, Boureau F, Brochet B, Bruxelle J, Cunin G, Fermanian J, Ginies P, Grun-Overdyking A, Jafari-Schluep H, Lanteri-Minet M, Laurent B, Mick G, Serrie A, Valade D, Vicaut E. Comparison of pain syndromes associated with nervous or somatic lesions and development of a new neuropathic pain diagnostic questionnaire (DN4). *Pain* 2005; 114: 29-36.
14. Freynhagen R, Baron R, Gockel U, Tolle TR. painDETECT: a new screening questionnaire to identify neuropathic components in patients with back pain. *Curr Med Res Opin* 2006; 22:1911-1920.
15. Bennett M. The LANSS Pain Scale: the Leeds assessment of neuropathic symptoms and signs. *Pain* 2001; 92:147-157.
16. Haanpaa M, Attal N, Backonja M, Baron R, Bennett M, Bouhassira D, Cruccu G, Hansson P, Haythornthwaite JA, Ianetti GD, Jensen TS, Kauppila T, Nurmikko TJ, Rice ASC, Rowbotham M, Serra J, Sommer C, Smith BH, Treede RD. NeuPSIG guidelines on neuropathic pain assessment. *Pain* 2011; 152:14-27.
17. Duong L, Cheriet F, Labelle H, Cheung KM, Abel MF, Newton PO, et al. Interobserver and intraobserver variability in the identification of the Lenke classification lumbar modifier in adolescent idiopathic scoliosis. *J Spinal Disord Tech* 2009; 22:448-455.
18. Lovelock CE, Anslow P, Molyneux AJ, Byrne JV, Kuker W, Pretorius PM, Coull A, Rothwell PM. Substantial observer variability in the differentiation between primary intracerebral hemorrhage and hemorrhagic transformation of infarction on CT brain imaging. *Stroke* 2009; 40:3763-3767.
19. Van Suijlekom HA, De Vet HC, Van Den Berg SG, Weber WE. Interobserver reliability in physical examination of the cervical spine in patients with headache. *Headache*. 2000; 40:581-586.
20. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005; 85:257-268.
21. Cohen J. A coefficient of agreement for nominal scales. *Educational and psychological measurement* 1960; XX:37-46.
22. Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. *Ann Acad Med Singapore* 1994; 23:129-138.
23. Treede RD, Jensen TS, Campbell JN, Cruccu G, Dostrovsky JO, Griffin JW, Hansson P, Hughes R. Nurmikko T, Serra J. Neuropathic pain: redefinition and a grading system for clinical and research purposes. *Neurology* 2008; 70:1630-1635.
24. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33:159-174.
25. Bron C, Franssen J, Wensing M, Oostendorp RA. Interrater reliability of palpation of myofascial trigger points in three shoulder muscles. *J Man Manip Ther*. 2007; 15:203-215.
26. Portney LGW, Watkins M.P. *Foundations of Clinical Research: Applications to Practice.* 2nd edition edn: Prentice Hall Health, Upper Saddle River, New jersey, USA, 2000.
27. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol* 1990; 43:551-558.
28. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990; 43:543-549.
29. Bhatia KS, Cho CC, Yuen YH, Rasalkar DD, King AD, Ahuja AT. Real-time qualitative ultrasound elastography of cervical lymph nodes in routine clinical practice: interobserver agreement and correlation with malignancy. *Ultrasound Med Biol* 2010; 36:1990-1997.
30. Richardson JK. The clinical identification of peripheral neuropathy among older persons. *Arch Phys Med Rehabil* 2002; 83:1553-1558.
31. Wald R, Bell CM, Nisenbaum R, Perrone S, Liangos O, Laupacis A, Jaber BL.Interobserver reliability of urine sediment interpretation. *Clin J Am Soc Nephrol* 2009; 4:567-571.
32. Mercadente S, Gebbia V, David F, Aielli F, Verna L, Casuccio A, Porzio G, Mangione S. Tools for identifying cancer pain of predominantly neuropathic origin and opioid responsiveness in cancer patients. *J Pain* 2009; 6:594-600
33. Steegers MA, Snik DM, Verhagen AF, van der Drift MA, Wilder-Smith OH.

Only half of the chronic pain after thoracic surgery shows a neuropathic component. *J Pain* 2008; 9:955-961.

34. Steegers MA, Wolters B, Evers AW, Strobbe L, Wilder-Smith OH. Effect of axillary lymph node dissection on prevalence and intensity of chronic and phantom pain after breast cancer surgery. *J Pain* 2008; 9:813-822.

35. Vadalouca A, Raptis E, Moka E, Zis P, Sykioti P, Siafaka I. Pharmacological treatment of neuropathic cancer pain: a comprehensive review of the current literature. *Pain Pract* 2012; 12:219-251.