

Aus dem Lehrstuhl für Neurophysiologie
der Medizinischen Fakultät Mannheim
Centrum für Biomedizin und Medizintechnik Mannheim
(Direktor: Prof. Dr. med. Rolf-Detlef Treede)

Optimizing the utility of Quantitative Sensory Testing for individual
diagnostics and development of a mechanism-based classification of
neuropathic pain.

Inauguraldissertation
zur Erlangung des Doctor scientiarum humanarum (Dr. sc. hum.)
der Medizinischen Fakultät Mannheim
der Ruprecht-Karls-Universität
zu
Heidelberg

vorgelegt von
Jan Vollert
aus Altena (Westf.)

2017

Dekan: Prof. Dr. med. Sergij Goerd

Referent: Prof. Dr. med. Rolf-Detlef Treede

CONTENTS

	Page
ABBREVIATIONS	1
1 INTRODUCTION	3
1.1 Pain and neuropathic pain.....	3
1.2 Peripheral sensory signal transduction.....	4
1.3 Nociception.....	4
1.4 Mechanisms of neuropathic pain.....	5
1.5 Sensory phenotyping.....	6
1.6 Aims and objectives.....	7
2 METHODS	10
2.1 Consortia and participating centers	10
2.2 Patient and participant selection.....	11
2.2.1 Patients.....	11
2.2.2 Healthy participants	12
2.2.3 Human surrogate models of neuropathic pain	12
2.3 QST protocol	13
2.4 z-transformation.....	14
2.5 Analysis of heterogeneity	15
2.6 Cluster analysis protocol	16
2.6.1 Validation sets	18
2.7 Individual sorting algorithm.....	18

2.7.1	Simplified phenotyping.....	20
2.7.2	Discrimination analysis against healthy participants	20
2.7.3	Effectiveness for treatment with oxcarbazepine	21
2.8	Sample size recommendations.....	22
2.9	Subgrouping human surrogate models.....	23
2.10	Patients in heuristic and mechanistic phenotypes	24
3	RESULTS.....	25
3.1	Patients and participants	25
3.1.1	Patients.....	25
3.1.2	Healthy participants	27
3.1.3	Human surrogate models	27
3.2	Analysis of heterogeneity	28
3.2.1	Healthy participants	30
3.2.2	Polyneuropathy.....	31
3.2.3	Peripheral nerve injury.....	31
3.3	Cluster analysis of patients.....	31
3.3.1	Sensory profiles of the three-cluster solution.....	33
3.3.2	Patient characteristics of the three clusters	35
3.3.3	Distribution of clusters across etiologies.....	39
3.4	Individual sorting algorithm.....	40
3.4.1	Sorting algorithm.....	40
3.4.2	Discrimination analysis against healthy participants	41
3.4.3	Deterministic and probabilistic algorithm	43
3.4.4	Effectiveness for treatment with oxcarbazepine	44
3.5	Frequency of phenotypes in clinical entities	45

3.5.1	Deterministic.....	47
3.5.2	Probabilistic.....	47
3.5.3	Simplified phenotyping.....	49
3.6	Sample size recommendations.....	49
3.7	Subgrouping human surrogate models.....	51
3.7.1	Cluster analysis of human surrogate models.....	55
3.7.2	Pattern-based sorting algorithm.....	56
3.7.3	Deterministic and probabilistic sorting.....	58
3.8	Patients in heuristic and mechanistic phenotypes.....	60
4	DISCUSSION.....	64
4.1	Heterogeneity between centers.....	65
4.2	Sensory phenotypes in patients.....	68
4.2.1	Sensory loss phenotype.....	69
4.2.2	Thermal hyperalgesia phenotype.....	70
4.2.3	Mechanical hyperalgesia phenotype.....	71
4.2.4	Sample size recommendations.....	71
4.3	Subgrouping human surrogate models.....	73
4.3.1	Nerve blocks as human surrogate model of denervation.....	73
4.3.2	Primary hyperalgesia as human surrogate model of peripheral sensitization.....	74
4.3.3	Secondary hyperalgesia as human surrogate model of central sensitization.....	74
4.3.4	Other human surrogate models.....	75
4.3.5	Mechanistic significance of heuristically found phenotypes.....	76
4.4	Limitations.....	77
4.5	Impact and conclusions.....	79

5 SUMMARY	81
6 REFERENCES	83
6.1 Publications that arose from the master thesis prefacing this work	96
6.2 Publications that arose from this work	96
7 LIST OF TABLES AND FIGURES	98
7.1 Tables.....	98
7.2 Figures	99
8 CURRICULUM VITAE	100
8.1 Publications	101
8.2 Oral presentations	103
8.3 Poster presentations.....	104
9 ACKNOWLEDGMENTS	106

ABBREVIATIONS

ARI	Adjusted Rand Index
AUC	Area Under Curve
AVI	Adjusted Variation of Information
BIC	Bayesian Information Criterion
BMBF	German Federal Ministry for Education and Research
CDT	Cold Detection Threshold
CI	Confidence Interval
CPT	Cold Pain Threshold
CRPS	Complex Regional Pain Syndrome
DB	Database
DET	Deterministic algorithm
DFNS	German Research Network on Neuropathic Pain
DMA	Dynamic Mechanical Allodynia
dPNP	diabetic Polyneuropathy
EM	Expectation Maximization
EMA	European Medicines Agency
H	Healthy sensory profile
HFS	electrical High-Frequency Stimulation
HIV	Human Immunodeficit Virus
HPT	Heat Pain Threshold
i.d.	intra-dermal
IASP	International Association for the Study of Pain
IENFD	Intra-Epidermal Nerve Fiber Density
IMI	Innovative Medicines Initiative
IN	Irritable Nociceptor

MDT	Mechanical Detection Threshold
MH	Mechanical Hyperalgesia phenotype
MPS	Mechanical Pain Sensitivity
MPT	Mechanical Pain Threshold
NB	Nerve Block mechanistic phenotype
NNT	Number Needed to Treat
NPSI	Neuropathic Pain Symptom Inventory
PH	Primary Hyperalgesia mechanistic phenotype
PHN	Post-Herpetic Neuralgia
PHS	Paradoxical Heat Sensation
PiNS	Pain in Neuropathy Study
PNI	Peripheral Nerve Injury
PNP	Polyneuropathy
PPT	Pressure Pain Threshold
QST	Quantitative Sensory Testing
RAD	Radiculopathy
ROC	Receiver Operating Characteristics
SH	Secondary Hyperalgesia mechanistic phenotype
SL	Sensory Loss phenotype
TH	Thermal Hyperalgesia phenotype
TN	Trigeminal Neuralgia
TRPV1	Transient Receptor Potential Vanilloid 1
TSL	Thermal Sensory Limen
UVB	Ultraviolet Radiation B
VDT	Vibration Detection Threshold
WDT	Warm Detection Threshold
WUR	Wind-Up Ratio

1 INTRODUCTION

1.1 Pain and neuropathic pain

Pain is defined as an “unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage” by the International Association for the Study of Pain (IASP) (Marskey et al., 1979). While acute pain serves as a warning and protective system activated by tissue damage or trauma, chronic and persistent pain may turn into a pathological state of its own (Loeser and Treede, 2008). If the pathophysiological basis is known, chronic pain can be subdivided into two major groups:

- Nociceptive pain is arising from activation of nociceptors (sensory neurons reporting actual or potential tissue damage). Chronic inflammation can lead to inflammatory pain, a chronic form of nociceptive pain. The nociceptive system itself is not directly affected, but may alter over time via sensitization (Loeser and Treede, 2008).
- Neuropathic pain results from an injury or disease to the nociceptive system (Treede et al., 2008; Finnerup et al., 2016) and can be caused peripherally (peripheral neuropathic pain) or in the central nervous system (central neuropathic pain).

Historically, peripheral neuropathic pain is classified based on the underlying disease or event initiating the nervous damage, such as diabetes, HIV, or chemotherapy-induced polyneuropathy (PNP), post-traumatic peripheral nerve injury (PNI), radiculopathy (RAD), trigeminal neuralgia (TN) or post-herpetic neuralgia (PHN) (Colloca et al., 2017). Treatment guidelines are often based on these etiologies, though it has become evident over the last decades that this approach is not sufficient, as first line treatment fails in over 50% of patients (Finnerup et al., 2015; Bouhassira and Attal, 2016). While written ten years ago, the following devastating statement remains largely valid: “*The management of patients with chronic NP [neuropathic pain, A/N] is complex and response to existing treatments is often inadequate. Even with well-established NP medications, effectiveness is unpredictable, dosing can be complicated, analgesic onset is delayed, and side effects are common*” (Dworkin et al., 2007).

The fact that common symptoms and signs of neuropathic pain appear across etiologies, while varying in pattern within etiologies, has led to the idea that distinct pathogenic mechanisms of neuropathic pain appear across etiologies (Fields et al., 1998; Woolf and Mannion, 1999; Campbell and Meyer, 2006; Baron et al., 2012; von Hehn et al., 2012). Treatment of neuropathic pain therefore should be aiming at mechanisms rather than at etiology.

1.2 Peripheral sensory signal transduction

The somatosensory (nervous) system comprises the afferent peripheral sensory receptor neurons, and their subsequent second-order neurons in the central nervous system (Kandel et al., 2013). All afferent sensory receptor neurons are pseudo-unipolar neurons based in the dorsal root (or trigeminal) ganglia with a single axon that divides at a T-junction into a peripheral axon, innervating skin or deep tissue, and a central axon, transmitting signals onto secondary neurons in the spinal cord or medulla oblongata (Squire et al., 2008). Sensory neurons differ, however, in degree of myelination, and (conclusively) signal conduction velocity (Kandel et al., 2013). Three major fiber types can be distinguished in primary afferent neurons:

- A β -fibers – thickly myelinated (nerve fiber diameter 6 – 12 μm), high conduction velocity (36 – 72 m/s) (Kandel et al., 2013) – innervate cutaneous mechanoreceptors (e.g. vibration, pressure) (Squire et al., 2008),
- A δ -fibers – thinly myelinated (nerve fiber diameter 1 – 6 μm), medium conduction velocity (3 – 36 m/s) (Kandel et al., 2013) – detect cold and noxious stimuli (Squire et al., 2008) and
- C-fibers – unmyelinated (nerve fiber diameter 0.2 – 1.5 μm), slow conduction velocity (0.4 – 2 m/s) (Kandel et al., 2013) – transmit warm and noxious stimuli (Squire et al., 2008).

1.3 Nociception

Noxious stimuli are transduced (e.g. via TRPV1) and subsequently transformed into action potential trains (e.g. via Nav 1.7, 1.8, etc.). These trains are conducted via small, thinly or unmyelinated A δ - or C-fibers. The divergent conduction velocity of these fibers results in the concept of first and second pain: the former, conducted by

faster A δ -fibers, is discriminative and rapidly induces efferent response, the latter, conducted by slower C-fibers, is less localized and of longer duration (Squire et al., 2008).

Nociceptors with A δ -fibers are sensitive to either mechanical stimuli and induce pain of sharp or pricking quality, or are sensitive to mechanical and heat stimuli and induce pain of burning quality (Kandel et al., 2013). Nociceptors innervated by C-fibers are polymodal or sensitive to mechanical and cold stimuli and induce pain of burning or freezing quality (Kandel et al., 2013). It should be noted that the majority of nociceptors are polymodal and respond to various stimuli, such as heat, cold, sharp or blunt pressure, or chemicals, but the sensitivity spectrum varies broadly between nociceptors (Kandel et al., 2013).

1.4 Mechanisms of neuropathic pain

According to a recent review (Colloca et al., 2017), along the nociceptive pathways, at three levels the generation of neuropathic pain can take place:

1. First-order nociceptor ion channels (peripheral sensitization).

Altered function of transduction channels (e.g. TRPV1 (Haanpaa and Treede, 2012)), as well as in- or decreased activity or expression of sodium, potassium and/or calcium channels in affected afferent nerves can induce spontaneous pain or hyperexcitability of the affected nerves, which has been shown for example in case of congenital overexpression of sodium channels, which can induce painful diseases like erythromelalgia (McDonnell et al., 2016). Similarly, increased sodium or decreased potassium channel function can induce hyperexcitable nociceptors (often called IN = irritable nociceptors) (Fields et al., 1998; Tesfaye et al., 2013). These can also cause spontaneous pain via ectopic activity. Increased pain sensitivity due to peripheral sensitization is limited to the site of injury, trauma or disease and called primary hyperalgesia (Treede et al., 1992; Hucho and Levine, 2007).

2. Second-order neurons (central sensitization).

Enhanced excitability of second order nociceptive neurons can increase their response to nociceptor input and widen their receptive field so that input from non-nociceptive sensory neurons induces nociceptive transduction (Woolf, 2011). Central sensitization can be induced by primary hyperalgesia and

accounts for secondary hyperalgesia (pain increased in spatial extend beyond the initial area of the injury or disease) and allodynia (painful sensation to non-painful stimuli) (Baron et al., 2013).

3. Inhibitory modulation.

Descending modulatory pathways and inhibitory interneurons can be impaired in patients with neuropathic pain, shifting the balance between pain inhibition and excitation further towards excitation (Colloca et al., 2017). It has been shown that the extent of conditioned pain modulation, where the perceived pain intensity of a steady painful test stimulus is reduced by applying a second tonic painful stimulus, is reduced in many patients with chronic pain (Lewis et al., 2012).

These mechanisms are, as stated above, present across etiologies of neuropathic pain (though varying in frequency), they may co-exist and enhance or facilitate each other, and, most importantly, are assumed to respond to distinct forms of treatment (Finnerup et al., 2015). Therefore, a mechanistic classification of neuropathic pain has been under debate for over 25 years now (Fields et al., 1998; Woolf et al., 1998; Baumgartner et al., 2002; Baron et al., 2012; von Hehn et al., 2012; Edwards et al., 2016).

1.5 *Sensory phenotyping*

To establish a mechanism-based classification, it is crucial to be able to detect and describe the mechanisms involved in the generation of pain in the individual patient. A first step can be patient-recorded questionnaires, capturing subjective reports of dimensions of neuropathic pain like the Neuropathic Pain Symptom Inventory (NPSI) (Bouhassira et al., 2004). As a second step, bedside testing of sensory signs like loss of thermal or mechanical detection or painful reaction to stimuli that are normally not perceived as painful may indicate involvement of mechanisms. However, both methods are subjective and hardly comparable between patients.

A comprehensive way to capture a patient's sensory function is Quantitative Sensory Testing (QST) (Krumova et al., 2012a). When performed in accordance with the DFNS (German Research Network on Neuropathic Pain) protocol (Rolke et al., 2006b), QST assess thermal and mechanical detection and pain thresholds, capturing various aspects of neuropathic pain:

1. Denervation/deafferentation – increased detection thresholds indicate loss of activity and/or presence of A β - (mechanical detection), A δ - (cold detection) or C-fiber sensory neurons.
2. Peripheral sensitization – heat hyperalgesia or increased deep pain sensitivity to blunt pressure in combination with unimpaired thermal detection may indicate irritable nociceptors.
3. Central sensitization – dynamic mechanical allodynia, increased pinprick pain sensitivity and cold hyperalgesia indicate central sensitization.
4. Modulatory descending pathways – an increased wind-up ratio (rating of a single painful stimulus compared to a series of ten such stimuli) indicates impaired descending noxious inhibition controls.

While QST has shown its capacity to identify and separate groups of patients with neuropathic pain (Maier et al., 2010; Smith et al., 2017), its usefulness for individual treatment is under discussion (Hansson et al., 2007; Backonja et al., 2013). Detection and pain thresholds vary broadly within healthy populations, and are influenced by age, gender, tested body region, and more problematic, by many factors that are impossible to control for, like genetics, epigenetics and individual development. Still, this is a frequent phenomenon in medicine and even more problematic in other tests assessing the nervous system (e.g. counting intraepidermal nerve fibers after skin biopsy (Isak et al., 2017)).

1.6 Aims and objectives

The European consortia IMI (Innovative Medicines Initiative) Europain, Neuropain and the DFNS have gathered QST data of 945 patients with peripheral neuropathic pain and 657 healthy participants with transient sensory changes due to surrogate models for neuropathic pain. In addition, reference data of healthy participants was collected in the DFNS (Rolke et al., 2006a; Blankenburg et al., 2010; Magerl et al., 2010; Pfau et al., 2014; Vollert et al., 2016a), and 188 healthy participants from additional German and European centers were included subsequently (Vollert et al., 2016a). All data have been collected in a central database in Bochum, Germany (Maier et al., 2010; Vollert et al., 2016a; Baron et al., 2017; Vollert et al., 2017b). While examination of patients and healthy participants was performed by physicians at the respective sites and collection and organization of data in a unified database

with high data quality was the central task of my master thesis (Vollert et al., 2015; Vollert et al., 2016b), scope of the present thesis was to adapt, develop and perform all mathematical analysis necessary to develop an individual algorithm to assign patients to sensory phenotypes that are linked to mechanisms of pain generation. In detail, aim of this work was to use this data to:

1. Perform a systematic analysis of heterogeneity of QST assessment of patients and healthy participants between the participating European centers, to show that comparability between centers is guaranteed, a central prerequisite for analyzing the data as a homogenous dataset.
2. Use unsupervised clustering methods to identify subgroups of sensory profiles appearing across etiologies of peripheral neuropathic pain and may indicate underlying mechanisms of pathophysiology.
3. Develop an algorithm that enables assignment of individual patients to one or more of the subgroups identified in (2) based on the patient's QST profile.
4. Apply the algorithm from (3) to 83 patients with peripheral neuropathic pain, whose pain relief after treatment with oxcarbazepine is known from a former study (Demant et al., 2014). Oxcarbazepine blocks sodium channels that are mainly located on small nerve fibers, and is therefore thought to be ineffective in patients with pain linked to deafferentation. In their study, Demant et al. found that patients with intact thermal detection show a significantly increased pain relief after treatment with oxcarbazepine compared to patients with loss of thermal detection. A pain relief that is significantly higher in a sensory subgroup from (2) with intact thermal detection compared to those with loss of thermal detection would indicate mechanistic variance in pain generation between subgroups.
5. Apply the algorithm from (3) to 335 patients with painful peripheral nerve injury, 151 patients with painful diabetic neuropathy, and 97 patients with post-herpetic neuralgia. Based on the frequency of each phenotype in each etiology, sample sizes of study populations that need to be screened to reach a sub-population large enough to conduct a phenotype-stratified study were to be calculated.
6. To create a database of human surrogate models studied with full QST profiles and to perform a similar cluster analysis and a similar individual algorithm as in (3) in 657 healthy participants with transient sensory changes

due to surrogate models for denervation, peripheral sensitization and central sensitization. These represent well-studied mechanisms.

7. To further validate the phenotypes identified in (2) by submitting them to the mechanism-based individual algorithm developed in (6). An emergence of similar phenotypes in surrogate models that are similar in mechanism would further strengthen the concept of sensory phenotypes representing mechanisms of neuropathic pain.

2 METHODS

2.1 *Consortia and participating centers*

The DFNS (<http://www.neuropathischer-schmerz.de>) was formed in 2002 and financed by a BMBF (federal ministry for education and research) grant to investigate mechanisms and treatments of neuropathic pain. Forming universities were the Ruhr University, Bochum, University of Schleswig Holstein, Kiel, Technical University, Munich, Ruprecht-Karls-University, Heidelberg, Johannes-Gutenberg-University, Mainz, University of Erlangen, University of Tübingen, University of Würzburg, University of Ulm. The study protocol was approved by the ethics committee of the University Hospital Kiel and subsequently by the ethics committees of all participating centers. Subsequently, the University Hospital of the Goethe-University, Frankfurt am Main, joined the DFNS and participated in collecting data from human surrogate models of neuropathic pain.

The EUROPAIN project (<http://www.imieuropain.org>) was founded in 2009 and financed by the European union's seventh framework programme. Data for this study were collected by the following centers: Aarhus University, Denmark, Ruhr-University, Bochum, Germany, University of Schleswig Holstein, Kiel, Germany, Technical University, Munich, Germany, Medical Faculty Mannheim, Ruprecht-Karls-University, Heidelberg, Germany. The ethics committee of each center approved the study protocol individually.

The NEUROPAIN project is an investigator-initiated project sponsored by Pfizer. Data for this study were collected by the following centers: Aarhus University, Denmark, Ruhr-University Bochum, Germany, University of Schleswig Holstein, Kiel, Germany, Technical University, Munich, Germany, Medical Faculty Mannheim, Ruprecht-Karls-University, Heidelberg, Germany, Université Versailles-Saint-Quentin, Versailles, France, Sapienza University, Rome, Italy, Helsinki University Central Hospital, Finland, Karolinska Institutet, Stockholm, Sweden, Benedictus Hospital Tutzing, Germany, Imperial College, London, United Kingdom, Neuroscience Technologies, Ltd., Barcelona, Spain. The ethics committee for each center approved the study protocol individually.

The Pain in Neuropathy Study (PiNS) was supported by the Wellcome Trust, and the European Union's Horizon 2020 research and innovation program. Data for this study were collected at the Oxford University, Oxford, United Kingdom, Imperial College, London, United Kingdom, and Sheffield Teaching Hospitals, Sheffield, United Kingdom. The ethics committee for each center approved the study protocol individually.

All participating centers underwent strict quality control (Geber et al., 2009; Magerl et al., 2010; Vollert et al., 2015) to ensure comparability of QST assessments between centers.

2.2 Patient and participant selection

2.2.1 Patients

[The following section has been taken in parts and modified from (Vollert et al., 2016b) and (Baron et al., 2017).]

Inclusion criteria for patients were carefully checked by a physician experienced in pain medicine at the local center. For each diagnosis, inclusion criteria were as follows:

- Polyneuropathy: pathological electroneurography or pathologically decreased vibration detection thresholds at two of four sites ($< 5/8$) at the lower limb, which could not be explained by another disease, or pain with polyneuropathy-type location and evidence of small fiber neuropathy based on skin punch biopsy, laser-evoked potentials or bedside thermal testing, which could not be explained by another disease.
- Peripheral nerve injury: history of traumatic nerve injury of the distal upper or lower limb and sensory-motor abnormalities confined to the innervation territory of the injured nervous structure or idiopathic sensory trigeminal neuropathy or iatrogenic mandibular neuropathy (i.e., inferior alveolar or lingual nerve neuropathy after various kinds of intraoral procedures) or trigeminal neuropathy secondary to compression, trigeminal neuropathy secondary to percutaneous lesions of the ganglion and sensory loss in the neuroanatomical adequate trigeminal territory.

- Post-herpetic neuralgia: unilateral zoster rash in the facial or thoracic area with post-zoster scarring, hypo- or hyperpigmentation in the affected dermatome or sensory deficit around the previous zoster rash determined by bedside-testing.
- radicular lesion: pain in the L5 and/or S1 dermatome and positive straight leg raising test or sensory deficit within the matching dermatome or diminished Achilles tendon reflex for S1 lesions and MRI of the lumbar spine confirming nerve root impairment by a herniated intervertebral disk or electromyography showing denervation in the L5 or S1 territory.

Exclusion criteria were age under 18 years, missing informed consent, insufficient language skills or other communication problems, pain treatment by topical local anesthetics for ≥ 7 days in the last 4 months or by topical capsaicin in the last 6 months, comorbidities treated by anticonvulsants or antidepressants, other pain locations with pain intensities ≥ 6 on ≥ 15 days/ month, other severe systemic or focal diseases of the central nervous system (e.g., stroke, spinal cord lesion), spinal canal stenosis, peripheral vascular disease (Fontaine stage II or higher), pending litigation and major cognitive or psychiatric disorders. In the cases of unilateral pain syndromes, contralateral neuropathies or painful conditions of the contralateral limb had to be excluded. Datasets were excluded in the case of incomplete records (e.g., no precise diagnosis available, more than 2 missing variables of the QST in the affected area, no information about age, gender or other demographic data).

2.2.2 Healthy participants

Healthy participants were included based on the recommendations by Gierthmühlen et al. (Gierthmühlen et al., 2015) and collected by ten centers from the DFNS, IMI Europain and Neuropain for quality assurance purposes (Vollert et al., 2015) during the certification process of these centers (Geber et al., 2009).

2.2.3 Human surrogate models of neuropathic pain

The following human surrogate models for neuropathic pain were conducted within the DFNS and included in the analysis (Klein et al., 2005; Vollert et al., 2017b):

- A-fiber-block (unpublished data collected in Kiel and Mannheim, methods as in (Ziegler et al., 1999)). Selectively blocking A-fibers leads to strongly decreased

sensory function (except warm detection, which is C-fiber mediated) and is a model for selective deafferentation/denervation of myelinated nerves (e.g. demyelinating polyneuropathy).

- Topical lidocaine cream application. Topical lidocaine induces loss of sensory function and therefore is a model of denervation/deafferentation (Krumova et al., 2012b).
- Topical capsaicin application, using cream, watery solution, or patch. The application of topical capsaicin induces peripheral sensitization, and can lead over time to a secondary hyperalgesia beyond the immediately affected area, a model for peripheral sensitization leading to central sensitization (Baron et al., 2013; Lotsch et al., 2015)
- UV-B light irradiation (unpublished data collected in Mannheim and published data, methods as in (Gustorff et al., 2013)). The sunburn model induces primary and secondary hyperalgesia similar to topical capsaicin (Gustorff et al., 2013)
- Intradermal capsaicin injection (unpublished data collected in Kiel, Mannheim and TU Munich, methods as in (Magerl and Treede, 2004)). The injection leads to secondary hyperalgesia, without inducing primary hyperalgesia first (Magerl and Treede, 2004).
- Cutaneous electrical high-frequency stimulation (HFS). Cutaneous HFS leads to mechanical hyperalgesia and allodynia, and therefore is a model of central sensitization (Lang et al., 2007).
- Muscular electrical high-frequency stimulation. HFS in the muscle is thought to lead to peripheral sensitization (Schilder et al., 2016).
- Topical menthol application. As topical menthol application leads to primary and secondary cold and mechanical hyperalgesia, it is a model for central sensitization (Wasner et al., 2004; Binder et al., 2011).
- Topical application of capsaicin solution in combination with lidocaine patch. This model induces primary hyperalgesia (peripheral sensitization) in combination with loss of function, e.g. numbness (Enax-Krumova et al., 2017).

2.3 QST protocol

QST in accordance with the DFNS protocol assesses 13 parameters: cold and warm detection thresholds (CDT and WDT, respectively), the thermal sensory limen during

alternating cold and warm stimuli and the number of paradoxical heat sensations (PHS) during this procedure, cold and heat pain thresholds (CPT and HPT, respectively), tactile (mechanical) detection threshold (MDT), mechanical pain threshold and sensitivity to pinprick stimuli (MPT, MPS), dynamical mechanical allodynia to touch with a brush, cotton wool or Q-tip (DMA), the wind-up ratio (WUR) of the perceived pain of a single pinprick stimuli compared to a series of ten stimuli, vibration detection threshold (VDT) and deep pain sensitivity to blunt pain (pressure pain threshold, PPT) (Rolke et al., 2006a).

Thermal sensory and pain thresholds were performed using either a TSA 2001-II (MEDOC, Israel) or a MSA (SOMEDIC, Sweden) that in- or decreased temperature by 1°C per second (Rolke et al., 2006b). For the TSA, six warm and cool stimuli were applied. The participant was asked whether he or she felt a cold or a warm stimulus, and the number of PHS (warm sensations during cold stimuli) was recorded. MDT was defined as the geometric mean of 5 series of stimuli ascending and descending between 0.25 and 512mN by a standardized set of von Frey hairs, mechanical pain threshold as the geometric mean of 5 series of stimuli ascending and descending by applying pinprick stimuli between 8 and 512mN (Rolke et al., 2006b). MPS and DMA were assessed by applying a total of 50 stimuli (35 pinprick and 15 light tactile in a balanced protocol) and asking patients to give a pain rating on a 0 (no pain) to 100 (worst pain imaginable) NRS scale. MPS was calculated as the geometric mean of the pain ratings of the pinprick stimuli, DMA as the geometric mean of the pain rating of the tactile stimuli. For the WUR the perceived intensity of a single pinprick stimulus was compared with that of a series of 10 repetitive pinprick stimuli of the same physical intensity on a 0-100 NRS scale, as an average of five series (Rolke et al., 2006b). VDT was assessed with a Rydel–Seiffer graded tuning fork (64 Hz, 8/8 scale, mean of three testing series) and PPT was determined over muscle with a pressure gauge device (FDN200, Wagner Instruments, USA), exerting forces up to 2000 kPa, as a mean of three series of ascending stimulus intensities, each slowly increasing (50 kPa/s) (Rolke et al., 2006b).

2.4 z-transformation

The initial assessment of 180 healthy participants for the DFNS reference database revealed that all parameters except PHS and DMA could be transformed (partly in

log-space) to a standard normal distribution (Rolke et al., 2006a; Magerl et al., 2010; Pfau et al., 2014). This process, called z-transformation, normalizes all values to a mean = 0 and a standard deviation = 1. Subsequently, all QST results of patients and healthy participants were transformed in accordance to this normalization. The z-transformation normalizes for age decade, gender and tested body region, thus making pain and detection thresholds comparable between patients with different age and gender, with nerves affected e.g. at the face or the feet. Abnormal values are defined as values beyond the 95% confidence interval. While individual z-scores are considered as abnormal if beyond ± 1.96 (Rolke et al., 2006a), for groups of patients, z-scores of ± 1.0 have been shown to be of clinical significance, as they include a relevant number of patients with abnormal values (Maier et al., 2010).

DMA and PHS, which do not normally occur in healthy participants, cannot be z-transformed. They are usually presented as percentages of presence, but for use in statistical analysis, they can be transformed to pseudo-z-values: DMA can be coded into a 0/2/3-variable representing no DMA (coded as 0), DMA with average pain ratings below 1 on a 0-100 NRS scale (coded as +2) and DMA with average pain ratings between 1 and 100 (coded as +3). PHS can be transformed to a binary 0/2-variable showing absence (coded as 0) or presence (coded as +2) of pathological values (Baron et al., 2017).

As many surrogate models were applied at the volar lower arm or upper thigh, which are not standardized areas, data of the subjects tested in the same area before treatment, or, if this data was unavailable, of the contralateral untreated side was compared used to calculate effect sizes (Cohen's *d* (Cohen, 2013)). This measure normalizes changes in the mean value before and after treatment to standard deviation, i.e. an effect size = 1 equals a change in the mean value after treatment that is equal to the standard deviation of the sample. There are no general interpretations of "good" effect sizes, but it is often considered that effect sizes below 0.3 can be considered small, and above 0.7 can be considered as large treatment effects (Cohen, 1992).

2.5 Analysis of heterogeneity

[The following section has been taken in parts and modified from (Vollert et al., 2016a).]

To assess the variability of QST results across the different centers, random effects models for the eleven normally distributed QST parameters were adopted, with the QST z-value of a certain parameter as the dependent variable and center as the random effect. For the analysis of cold detection thresholds in patients suffering from polyneuropathy, let Y_{ij} be the CDT z-value of j th patient ($1 \leq j \leq N$) in the i th center ($1 \leq i \leq M$), the model is specified as:

$$F(1): Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

$$F(2a): \varepsilon_{ij} \sim N(0, \sigma^2)$$

$$F(2b): \alpha_i \sim N(0, \sigma_\mu^2)$$

where μ is the overall mean CDT z-value across all centers, α_i is the unobserved center-specific random effect and ε_{ij} is the individual unexplained effect. The model is fitted so that ε_{ij} and α_i are normally distributed with a mean value = 0 (F(2a) and F(2b)). The estimated center-specific mean z-values ($\mu + \alpha_i$ for $1 \leq i \leq M$) and the corresponding 95% confidence intervals from Model (1) are graphically displayed in forest plots separately for PNP, PNI and healthy participants, with the Y-axis showing a running number for the centers for all 11 QST parameters and the X-axis presenting the center-specific mean z-value. The I^2 statistic (Higgins and Thompson, 2002) was employed to quantify the heterogeneity of mean z-values across the centers. This statistic ranges from 0 to 100% and it measures the percentage of variation across studies that is due to true heterogeneity rather than chance. The confidence interval estimates of I^2 were computed based on the test-based method of Higgins and Thompson (Higgins and Thompson, 2002).

2.6 Cluster analysis protocol

[The following section has been taken in parts and modified from (Baron et al., 2017).]

A cluster analysis was performed to unravel distinguishable subgroups of QST-profiles. The procedure of this analysis was identical for patients suffering from peripheral neuropathic pain (Baron et al., 2017) and healthy participants under conditions that induce human surrogate models of neuropathic pain, in order to

guarantee comparability between the results. We made no a-priori-assumptions about the expected number of clusters and applied a k-means-approach for a number of k clusters ranging from 2 to 10 (MacQueen, 1967). A two-step protocol was used to determine the optimum number of clusters:

- Negative silhouettes as exclusion criterion. A silhouette width is a value that can be calculated for each subject in each cluster and ranges between -1 and +1 (Rousseeuw, 1987). Positive values near +1 indicate a subject that can unambiguously be allocated to a certain cluster, values near zero indicate subjects that are on the edge between clusters, and negative values indicate that these subjects are allocated to a cluster that is not their nearest cluster. This is possible because if they would be allocated to their nearest cluster, the cluster mean would shift so that other subjects are then in a cluster that is not their nearest. A mean silhouette width well above zero for a cluster indicates that said cluster is clearly separated from the others, while a mean silhouette width below zero indicates a cluster that is rather a statistical artifact than a real subgroup. Therefore, a high count of negative silhouettes or a cluster with a mean silhouette width below zero indicate a cluster solution that is highly fragmented. Thus, to control for statistical artefacts, we excluded all solutions with at least one cluster with a negative mean silhouette width, and all solutions with over 10% negative silhouette widths.
- Comparability between clustering methods. The remaining solutions were compared to two additional cluster methods with significantly different mathematical background, validating that the final solution is not dependent on the clustering method. We decided on a robust hierarchical clustering method (maximum linkage) and an Expectation Maximization (EM) algorithm (Dempster et al., 1977). Both solutions were compared to the initial k-means clustering, using the adjusted rand index (ARI) and the adjusted variation of information (AVI). While the ARI measures similarity on a scale from 0 – 1 (high values are preferable), the AVI measures dissimilarity on the same scale (low values are preferable) (Rand, 1971). Final criterion for the decision between otherwise equal cluster solutions was the Bayesian Information Criterion (BIC) which captures the gain of information by increased number of clusters. As a rule of thumb, the higher number of clusters is preferable, if the

difference between the BICs of both solutions (Delta-BIC) is larger than ten (Schwarz, 1978).

2.6.1 Validation sets

To investigate the stability of the findings, for both cluster analyses (patients and surrogate models) a validation set was formed, in which additional k-means cluster analyses were performed, with a fixed number of clusters (as defined above).

For patients suffering from peripheral neuropathic pain, this validation set was formed from patients with PNP, PNI and PHN who were collected either within the DFNS after closure of the initial database (Maier et al., 2010) or within the European consortium for treatment studies with oxcarbazepine and lidocaine (Demant et al., 2014; Demant et al., 2015).

For human surrogate models of neuropathic pain, no similar cohort was available. Therefore, 50% of all data (chosen by random number assignment) were excluded from the initial analysis to form the validation set.

2.7 Individual sorting algorithm

[The following section has been taken in parts and modified from (Vollert et al., 2017a).]

As QST z-values are (per definition) normally distributed, our approach was based on normally distributed probabilities. For each QST z-value of each parameter i and patient n , a probability can be calculated for a phenotype to be present based on the density function of said phenotype:

$$F(3): p_{i,n,m} = \frac{1}{\sqrt{2\pi\sigma_{i_m}^2}} \exp\left(-\left(\frac{(x_{in}-\mu_{i_m})^2}{\sqrt{2\sigma_{i_m}^2}}\right)\right)$$

With i = one of 13 QST parameters, n = the n th patient in a set of patients, m = one of the phenotypes determined in the cluster analysis of patients described above and conclusively, σ_{i_m} being the standard deviation of the i th QST parameter for the m th phenotype in the defining dataset (Baron et al., 2017), μ_{i_m} being the mean z-value of

the same QST parameter and phenotype in the defining dataset (Baron et al., 2017), and finally x_{i_n} being the z-value found in the n th patient for the i th QST parameter.

While this function will always reach its maximum at $x_{i_n} = \mu_{i_m}$, in relation to broadness of the standard deviation, density functions can become broader or narrower. This affects the maximum value the density function can reach. To control for these more or less broad functions, we normalized the formula so that a value that is equal to the mean of the phenotype equals 100%, leading to

$$F(4): p_{i_n,m}^* = \frac{1}{\sqrt{2\pi\sigma_{i_m}^2}} \exp\left(-\left(\frac{(x_{i_n}-\mu_{i_m})^2}{2\sigma_{i_m}^2}\right)\right) / \frac{1}{\sqrt{2\pi\sigma_{i_m}^2}} \exp\left(-\left(\frac{(\mu_{i_m}-\mu_{i_m})^2}{2\sigma_{i_m}^2}\right)\right)$$

which can be simplified to

$$F(5): p_{i_n,m}^* = \exp\left(-\left(\frac{(x_{i_n}-\mu_{i_m})^2}{2\sigma_{i_m}^2}\right)\right)$$

The resulting probability value is ranging from 0% to 100% and can be calculated for all $i = 13$ QST parameters and m phenotypes. By averaging the probability over the 13 QST parameters, we quantify the similarity of the individual patient's QST profile to the mean profile of each of the phenotypes.

As a simple way of categorizing patients into phenotypes, we suggest sorting each patient to the phenotype with the highest probability value:

1. Calculate F(5) for each of the 13 QST parameters. Use μ and σ for phenotype 1.
2. Average the 13 probabilities. The resulting value is the probability for this patient to show this phenotype.
3. Repeat steps 1 and 2, using μ and σ for the next phenotype. Repeat for all phenotypes.
4. Allocate the patient to the phenotype with the highest probability value.

The algorithm as described above was applied to the patients from the original cohort (Baron et al., 2017) to demonstrate its general sorting capacity.

2.7.1 *Simplified phenotyping*

As the DFNS protocol is comprehensive, it might be too complex to be applied in all clinical settings and in large clinical trials. If the cluster analysis showed that few parameters explain large parts of the variance between the discovered phenotypes, the accuracy of a phenotyping based on these in comparison to a phenotyping using the full protocol is analyzed.

2.7.2 *Discrimination analysis against healthy participants*

To show if and how the algorithm can discriminate patients with neuropathic pain from healthy participants, we introduced a fourth probability - not for a phenotype, but for being healthy. For this purpose, we applied the definition of QST z-values, to which a group of healthy participants ideally has a z-value mean = 0 (μ) with a standard deviation = 1 (σ) for each QST parameter. The original cluster patient cohort (Baron et al., 2017) ($n = 902$) and $n = 188$ healthy participants from the European cohort (Vollert et al., 2016a) underwent a modified version of the algorithm:

1. Calculate $F(5)$ for each of the 13 QST parameters. Use μ and σ for healthy.
2. Average the 13 probabilities. The resulting value is the probability for this patient to show a healthy profile.
3. Repeat steps 1 and 2, using μ and σ for all phenotypes.

As this version of the algorithm does not sort each subject simply to the phenotype with the highest probability, this leaves every subject with a series of probabilities, one for each of the phenotypes of neuropathic pain, and one for being healthy.

The probability of being healthy was used for a Receiver Operating Characteristics (ROC) plot (Zweig and Campbell, 1993). This graphical tool for assessing discriminatory power plots the false-positive rate (1 - specificity) on the x-axis versus the sensitivity of detecting patients on the y-axis for all possible probability values of being healthy. Each step in the ROC plot represents the specificity and sensitivity of one certain percentage. To assess the overall quality of separating healthy participants and patients via the probability for being healthy, the area under curve (AUC) and its 95% confidence interval were calculated (DeLong et al., 1988). To define a minimum probability, at which a subject should be considered being healthy,

the probability with the highest Youden-Index (sensitivity minus false-positive rate (Youden, 1950)) was chosen.

2.7.3 Effectiveness for treatment with oxcarbazepine

It has been shown in a recent study that the effectiveness of oxcarbazepine as treatment for neuropathic pain is dependent on the sensory phenotype of the patients (Demant et al., 2014). Oxcarbazepine is a blocker of voltage-gated sodium channels (McLean et al., 1994) and therefore has the potential to reduce neuropathic pain caused by overexpression or increased sensitivity of sodium channels (Ichikawa et al., 2001). This effect is, however, only possible in patients where this specific mechanism is present, and oxcarbazepine will therefore not help patients whose pain is generated more centrally (Katz et al., 2008). In their paper, Demant et al. phenotyped the patients as “irritable nociceptor” (IN, intact thermal detection, thermal or mechanical hyperalgesia) and “non-irritable nociceptor” (the remainder) (Demant et al., 2014). It was found that pain reduction was only significant in patients with “irritable nociceptor” phenotype. The number-needed-to-treat (number of patients that has to be treated with oxcarbazepine to find at least one patient with a pain reduction of at least 50% (Tramer and Walder, 2005)) was found to be 3.9 (95% CI: 2.3 – 11.5) for irritable nociceptor and 13.0 (95% CI: 5.2 - ∞) for non-irritable nociceptor. Using the original data provided by the principal investigators, the cohort underwent the individual algorithm to determine each patient’s cluster-based sensory phenotype. Based on mechanistic assumptions for the phenotypes identified in the previous steps, a hypothesis was formed which phenotype would respond to oxcarbazepine treatment. The treatment outcome was compared between cluster-based phenotypes and irritable/non-irritable nociceptor phenotyping, to analyze whether cluster-based phenotyping is similarly effective as (and non-inferior to) irritable/non-irritable nociceptor phenotyping as predictor for effectiveness of oxcarbazepine.

Three metrics were applied to show cluster-based phenotype-specificity of oxcarbazepine-related pain relief and to compare treatment prediction between phenotyping methods (cluster-based vs. irritable/non-irritable nociceptor):

- Treatment-phenotype interaction in a mixed effects model with pain relief as dependent variable, treatment (placebo/verum) and phenotype as fixed effects, patient as random effect and baseline pain as covariate, similar to the

model in the published study (Demant et al., 2014). This model was tested for both deterministic and probabilistic phenotyping.

- Mean pain reduction in the verum phase. The pain reduction between baseline and week six in the verum phase for IN patients and patients with a cluster-based phenotype that indicates effectiveness of oxcarbazepine was compared (two-tailed t-test). Cluster-based phenotyping was considered effective in case of a non-significant test or a significant test in combination with higher mean pain relief in the cluster-based phenotype compared to IN (non-inferiority).
- Number-needed-to-treat. The NNT was calculated for the cluster-based phenotype along with its 95% confidence interval (Tramer and Walder, 2005). Cluster-based phenotyping was considered effective in case of a lower NNT in the cluster-based phenotype compared to IN or a higher NNT with CIs overlapping between the NNTs for IN and cluster-based phenotyping (non-inferiority).

2.8 *Sample size recommendations*

[The following section has been taken in parts and modified from (Vollert et al., 2017a).]

If a new drug would be tested for efficacy in a phenotype-stratified subgroup with neuropathic pain of any single etiology, this would only be feasible if said phenotype appears in a relevant frequency within this etiology. To show how frequent these phenotypes are across three common etiologies of neuropathic pain, we applied the algorithm to patients suffering from neuropathic pain due to diabetic polyneuropathy, peripheral nerve injury or post-herpetic neuralgia from the databases of our previous studies (Maier et al., 2010; Demant et al., 2014; Demant et al., 2015; Themistocleous et al., 2016; Baron et al., 2017).

Based on the frequencies found in the clinical entities, we calculated the size of a group of patients that need to be screened with either full or simplified phenotyping to find a sub-population large enough to perform a trial that still reaches a power of 80% for an effect size of 0.3, 0.5 and 0.7 at an alpha-level of 0.05, for a crossover and parallel design. The sample sizes presented in this thesis are examples and can be tailored to the needs of any planned RCT. We recommend the usage of the free software G*Power (Faul et al., 2007), but many other statistical packages provide

similar tools. The following information is required before starting: alpha-level (usually 0.05), power (usually 0.8, 0.9 or 0.95), test family (usually t-test for independent (parallel design) or dependent (crossover design) means, or chi-squared for dichotomous outcome), and the estimated effect size in the phenotype of interest. Effect sizes are related to mean treatment effect and standard deviation between treatment response, e.g., a mean effect of 2 on a 0-10 NRS scale with a standard deviation of 4 corresponds to an effect size 0.5, a mean effect of 3.5 with a standard deviation of 5 an effect size of 0.7, and a mean effect of 1 with a standard deviation of 3 corresponds to an effect size of 0.3, and many other combinations are possible. With this information, the size of the subgroup of patients with the phenotype of interest that needs to be included can be calculated. To determine the size of the overall population which needs to be screened to find a subgroup of the calculated size, divide the subgroup size by the frequency of the phenotype in the etiology of interest as presented in the results section, in regard to the algorithm used (deterministic / probabilistic) and the phenotyping protocol (full / simplified).

2.9 *Subgrouping human surrogate models*

Human surrogate models were analyzed for patterns using a two-way strategy:

1. A hypothesis-free cluster analysis was applied, using the protocol as defined for the cluster analysis in patients suffering from neuropathic pain (see 2.6).
2. Since, unlike in the analysis of patients, for some human surrogate models, mechanisms are clearly described, as an additional means of subgrouping, a pattern-based individual algorithm using the method as developed based on patients suffering from neuropathic pain (see 2.7) was applied. This algorithm was in this case based on six surrogate models with a clearly described mechanism (A-fiber block and topical lidocaine for nerve block, topical capsaicin or UVB radiation for primary hyperalgesia, and i.d. capsaicin injection and cutaneous HFS for secondary hyperalgesia).

2.10 Patients in heuristic and mechanistic phenotypes

To analyze similarities between the mechanistic sensory subgroups described above and the sensory phenotypes found by hypothesis-free pattern searching methods in patients suffering from neuropathic pain, the patients suffering from neuropathic pain underwent the algorithm from 2.9 and the result was compared to each patient's sensory phenotype as determined using the algorithm developed for patients in 2.7.

Agreement between algorithms was assessed using Cohen's Kappa (Fleiss et al., 2003). We then applied a probabilistic sorting algorithm to estimate the prevalence of the three presumed mechanisms in this cohort of patients; this allows each patient to be assigned to more than one mechanism.

3 RESULTS

3.1 *Patients and participants*

Patients and participants for the various analyses were not identical, and are therefore described briefly below.

3.1.1 *Patients*

Essential basis of all analyses of patients was the European cohort, formed for the cluster analysis in patients (Baron et al., 2017). This cohort comprises $n = 902$ patients with peripheral neuropathic pain. Formation of this cohort is shown in Figure 1.

For the analysis of heterogeneity, a sub-collective of this cohort was formed: only data from centers which provided at least ten patients with painful polyneuropathy and ten patients with painful peripheral nerve injury were included. These ten centers provided 217 patients with painful PNP and 150 patients with painful PNI.

For the validation set of the cluster analysis in patients, a second cohort was formed from patients that were included by the DFNS after the closure of the initial database (Maier et al., 2010) and in the IMI for pharmacological studies (Demant et al., 2014; Demant et al., 2015). This group comprised $n = 233$ patients with peripheral neuropathic pain due to polyneuropathy, peripheral nerve injury or post-herpetic neuralgia.

For estimating frequency of phenotypes in etiologies of neuropathic pain and suggesting sample sizes for phenotype-stratified studies, the European cohort ($n = 902$) and the validation cohort ($n = 233$) were merged with data from the PiNS cohort ($n = 209$). From the resulting patient group, all patients with painful peripheral nerve injury ($n = 335$), painful diabetic polyneuropathy ($n = 151$), and painful post-herpetic neuralgia ($n = 97$) were extracted.

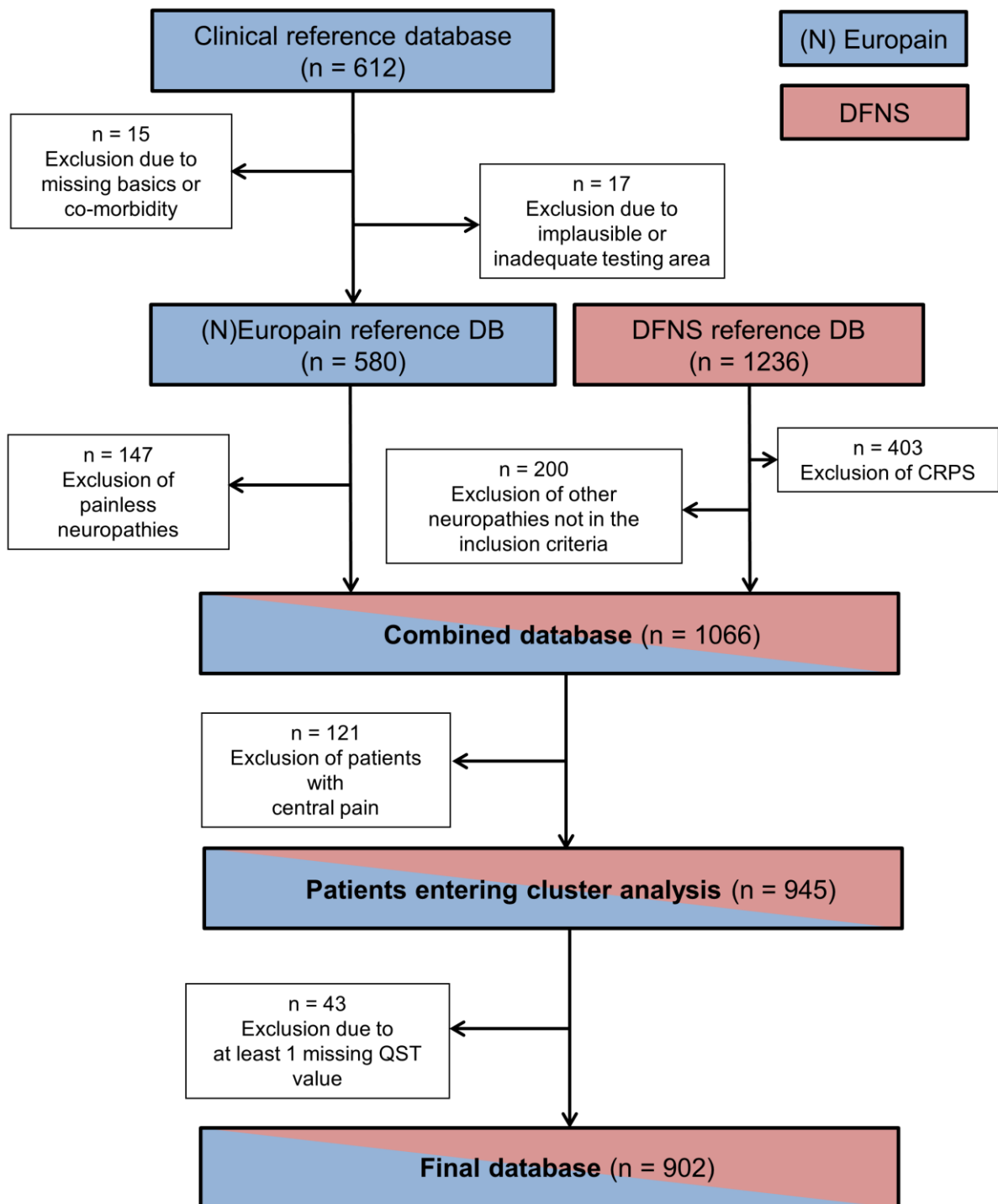


Figure 1: CONSORT flowchart of data inclusion for the European cohort. CRPS: Complex Regional Pain Syndrome, DB: database. From (Baron et al., 2017).

3.1.2 Healthy participants

For the analysis of heterogeneity and for separation of patients and healthy participants in the individual sorting algorithm, QST data from $n = 188$ healthy participants provided for ensuring quality control (Vollert et al., 2015) by ten European centers was included (Vollert et al., 2016a).

3.1.3 Human surrogate models

Many participants were tested more than once under various models, so the number of participants is less meaningful than the number of QSTs under surrogate models, which will therefore be referred to subsequently. A total of $n = 657$ QSTs under surrogate models could be included (see Table 1).

model	n	Gender: n (%) female	Age: mean (range)	participating centers
nerve block				
A fiber block	24	12 (50%)	25 (21 - 39)	Kiel, Mannheim
Topical lidocaine	41	20 (49%)	34 (19 - 69)	Bochum
primary hyperalgesia				
Topical capsaicin	273	147 (53%)	25 (15 - 75)	Bochum, Frankfurt, Mannheim
UVB	158	51 (32%)	24 (19 - 42)	Frankfurt, Mannheim
secondary hyperalgesia				
Capsaicin injection	36	19 (53%)	32 (23 - 68)	Kiel, Mannheim, Munich
Cutaneous HFS	12	3 (25%)	36 (24 - 57)	Mannheim
mixed				
Topical capsaicin, secondary hyperalgesia	37	15 (41%)	24 (19 - 39)	Mannheim
UVB, secondary hyperalgesia	22	0 (0%)	24 (24 - 24)	Mannheim
Muscular HFS	15	7 (47%)	24 (19 - 27)	Mannheim
Topical menthol	11	0 (0%)	25 (23 - 28)	Kiel
Topical lidocaine + topical capsaicin	28	17 (61%)	30 (20 - 75)	Bochum

Table 1: Participants under human surrogate models of neuropathic pain included in the analysis. HFS: High Frequency Stimulation, UVB: Ultraviolet Radiation B. From (Vollert et al., 2017b).

3.2 *Analysis of heterogeneity*

[The following section has been taken in parts and modified from (Vollert et al., 2016a).]

The forest plots for healthy participants, patients with polyneuropathy and peripheral nerve injury are shown in Figure 2 for each of the 11 normally distributed QST parameters, each of the 10 centers and mean effect. The forest plots show for each center, QST parameter and separately for healthy participants, PNP and PNI a center specific mean of each QST parameter, along with its 95% confidence interval. These confidence intervals indicate how individual and center-specific variance relate to each other: broader confidence intervals indicate high variance within centers that is not similar between all centers, narrow confidence intervals indicate high homogeneity within centers or variance that is similar between all centers.

The corresponding I^2 values and their confidence intervals can be found in Table 2.

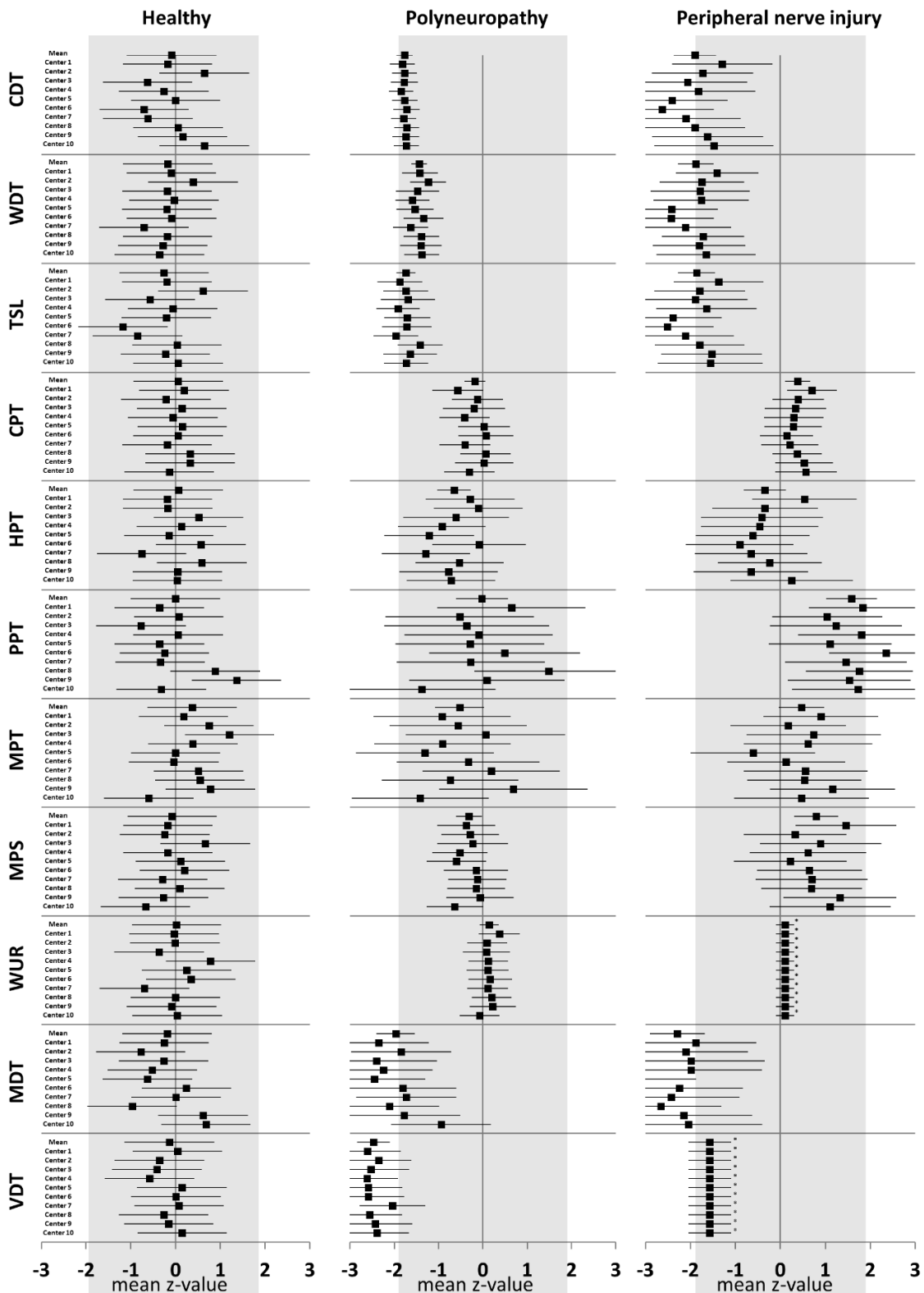


Figure 2: Forest plot of center-specific and overall mean QST z-values with 95% confidence intervals for healthy participants (left graph), patients with polyneuropathy (middle graph) and peripheral nerve injury (right graph). Note that confidence intervals are not equal to standard deviation or standard error of the data of each center. *model identified no center-specific mean, therefore only the overall mean and confidence interval is shown. All values represent the overall mean with 95% confidence interval. From (Vollert et al., 2016a).

3.2.1 Healthy participants

All center-specific mean z-values were found to be well within the normal range of 0 ± 1.96 , the mean z-values per parameter are within 0 ± 0.20 , with two exceptions: TSL (-0.26, 95% CI: -0.62; 0.10) and MPT (0.37, 95% CI: -0.01; 0.75). The overall mean z-value of all parameters and centers was found to be *exactly* 0.00. The 95% confidence intervals varied in width between 0.5 and 1.4, with a mean of 0.9. Broad confidence intervals indicate a high degree of individual variance within centers that is not found between all centers (i.e. the center-specific mean is uncertain). Inclusion of more participants per center might reduce the influence of this effect. In the given data, the largest part of variance is of individual origin, not between centers, which is represented in the I^2 values in Table 2 (first column), showing a heterogeneity of 0% for all parameters except PPT (41.8%, 95% CI: 0.0%; 66.1%) and MDT (5.4%, 95% CI: 0.0%; 18.1%).

	Healthy participants	Polyneuropathy	Peripheral nerve injury
CDT	0.0% (0.0% - 57.0%)	0.0% (0.0% - 0.0%)	0.0% (0.0% - 73.5%)
WDT	0.0% (0.0% - 99.9%)	0.0% (0.0% - 54.0%)	0.0% (0.0% - 71.1%)
TSL	0.0% (0.0% - 50.8%)	0.0% (0.0% - 57.9%)	0.0% (0.0% - 77.0%)
CPT	0.0% (0.0% - 80.0%)	0.0% (0.0% - 91.1%)	0.0% (0.0% - 50.6%)
HPT	0.0% (0.0% - 100.0%)	0.0% (0.0% - 98.9%)	0.0% (0.0% - 79.0%)
PPT	41.8% (0.0% - 66.1%)	0.0% (0.0% - 82.3%)	0.0% (0.0% - 58.2%)
MPT	0.0% (0.0% - 76.5%)	0.0% (0.0% - 99.8%)	0.0% (0.0% - 76.8%)
MPS	0.0% (0.0% - 100.0%)	0.0% (0.0% - 59.3%)	0.0% (0.0% - 69.2%)
WUR	0.0% (0.0% - 45.3%)	0.0% (0.0% - 33.5%)	0.0% (0.0% - 0.0%)
MDT	5.4% (0.0% - 18.1%)	0.0% (0.0% - 92.1%)	0.0% (0.0% - 56.4%)
VDT	0.0% (0.0% - 89.1%)	0.0% (0.0% - 26.7%)	0.0% (0.0% - 0.0%)

Table 2: I^2 Index of heterogeneity, ranging from 0% (no heterogeneity) to 100% (perfect heterogeneity) for all eleven normally distributed QST parameters for healthy participants ($n = 188$), polyneuropathy patients ($n = 217$) and patients with peripheral nerve injury ($n = 150$). In brackets: lower and upper boundary of the 95% confidence interval. All values except PPT and MDT in healthy participants are zero, though the confidence intervals often cover a broad range up to 100% due to the broad confidence intervals of center specific mean z-values. From (Vollert et al., 2016a).

3.2.2 *Polyneuropathy*

Center specific mean z-values for detection thresholds were characterized by loss of thermal and mechanical detection (CDT, WDT, TSL, MDT, VDT), which is typical for patients suffering from polyneuropathy. Note that the small confidence intervals (esp. for cold and warm detection threshold) do not reflect little variation in the dataset, but rather that only a small part of this variation is assigned to the center, the remainder is assigned to individual effects (ε_{ij} in $F(1)$) that appear across centers in similar form. For pain thresholds, ranges of center specific mean z-values and confidence intervals were found to be much broader, especially for MPT and PPT. Mean z-values scattered within the normal range of 0 ± 1.96 , but very broad confidence intervals show that this indicates merely high individual variance between PNP patients rather than systematic deviation by single centers. This is represented in I^2 values indicating heterogeneity of 0% between the centers for all parameters (Table 2, middle column).

3.2.3 *Peripheral nerve injury*

Similar as for PNP, center specific mean z-values for detection thresholds in PNI patients showed loss of function. Pain thresholds are mainly decreased (with HPT as exception), as expected for PNI patients, who often show a combination of loss of detection and gain of nociception. Confidence intervals are broader compared to PNP, indicating that a larger amount of variance is found within instead of between centers. For wind-up ratio and vibration detection threshold, individual variation was found to superimpose any possible center specific effects (a_i in $F(1)$) in patients suffering from PNI, so the forest plot (Figure 2, right column) presents only the overall mean and its confidence interval instead of center specific means for all centers (μ instead of $\mu + a_i$). As almost all center specific mean z-values lie within the confidence intervals of each other, the I^2 index of heterogeneity is 0% for all parameters.

3.3 *Cluster analysis of patients*

[The following section has been taken in parts and modified from (Baron et al., 2017).]

According to the frequency of negative silhouette widths we excluded the solutions with 4 - 10 clusters because they each presented at least one cluster with a negative mean silhouette width that indicated an artifact. Furthermore, in each of these solutions negative silhouettes were frequent (15 - 23%). The remaining two and three cluster solutions were compared with two additional, mathematically different clustering algorithms for the same number of clusters. Compared with agglomerative hierarchical cluster analysis, both 2 and 3-cluster solutions were equal according to the ARI criterion, but the three-cluster solution was better according to the AVI criterion. In comparison to the EM algorithm, the two-cluster solution failed to show similarity between k-means and EM clustering (ARI almost zero, AVI almost 1). Since the Delta-BIC also strongly preferred the three-cluster solution (Table 3), the three-cluster solution was used for further analysis as the optimal number of clusters. The replication data set was also subjected to a k-means cluster analysis with $k = 3$.

n (cluster)	silhouette width:			comparison to hierarchical:			comparison to EM:	
	mean °	minimum mean per cluster°°	negative (%)°°°	ARI*	AVI**	ARI*	AVI**	BIC***
2	0.29	0.28	0.7%	0.30	0.67	0.01	0.95	0
3	0.25	0.13	4.8%	0.30	0.56	0.22	0.69	708
4	0.23	-0.24	14.5%					
5	0.20	-0.28	16.4%					
6	0.15	-0.10	22.6%					
7	0.17	-0.07	21.2%					
8	0.19	-0.003	16.3%					
9	0.19	-0.02	16.0%					
10	0.21	-0.06	14.7%					

Table 3: Decision on the number of clusters. Green: optimum number of clusters according to this criterion. °Mean silhouette width per cluster. A value below zero indicates clusters that do not separate from other clusters. Measure of discriminatory power (0 to 1). 0: no discrimination, 1: perfectly separated clusters (high values are preferred) °°Measure of fragmentation of solution (-1 to +1). -1: cluster that is solely a fragment, +1: a solution that is not fragmented (solutions with values below zero were discarded (yellow)) °°°Measure of fragmentation of solution (0% to 100%). 0%: no fragmentation, 100% a completely fragmented solution (solutions with values above 10% were discarded (yellow)) *ARI (Adjusted Rand Index): Measure of similarity (0 to 1). 0: only random identity, 1: perfect identity (high values are preferred) **AVI (Adjusted Variation of Information): Measure of dissimilarity (0 to 1). 0: no dissimilarity, 1: strong dissimilarity (low values are preferred) ***Delta-BIC (Bayesian Information Criterion): Measure of gain of information by increasing cluster number. If delta-BIC > 10, the higher cluster number is recommended. Modified from (Baron et al., 2017).

3.3.1 Sensory profiles of the three-cluster solution

Figure 3 shows the mean z-score sensory profiles for the test data set (Fig. 3A) and the validation data set (Fig. 3B). In both data sets, the clusters represented similar percentages of patients: cluster 1 was the largest (42% in A, 53% in B), followed by cluster 2 (33% in A and B) and cluster 3 (24% in A, 14% in B). Sensory profiles were also replicated excellently. For non-nociceptive temperature sensation (CDT, WDT, TSL), clusters 1 and 3 exhibited pronounced deficits with mean z-scores near -2, while temperature sensation was essentially normal in cluster 2. This offset was similar for thermal pain sensitivity (CPT, HPT), but here clusters 1 and 3 exhibited less of a deficit, while cluster 2 exhibited significant sensory gain. Cluster 2 was therefore given the label "thermal hyperalgesia". For mechanical pain sensitivity

(PPT, MPT, MPS), the rank order between clusters was different and cluster 1 and 3 were separated: while there was again a deficit for cluster 1, cluster 3 exhibited significant sensory gain. Cluster 3 was therefore given the label "mechanical hyperalgesia". Wind-up did not differentiate between clusters. For non-nociceptive touch sensation (MDT, VDT), cluster 2 was again close to normal, cluster 3 had some deficit, and cluster 1 exhibited the most pronounced deficit. Cluster 1 was given the label "sensory loss", because it was characterized by negative mean z-scores across all QST parameters. Dynamic mechanical allodynia (DMA) was most pronounced in cluster 3, which also exhibits the most pronounced hyperalgesia to pinprick (MPT, MPS) and blunt pressure (PPT). Paradoxical heat sensations were most pronounced in cluster 1, associated with diminished cold detection (CDT) but not cold hyperalgesia (CPT).

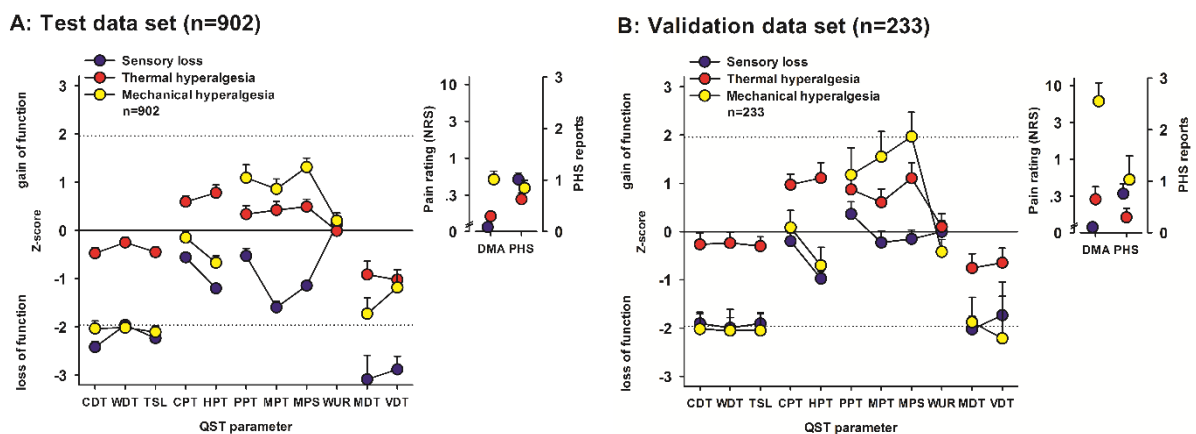


Figure 3: Sensory profiles of the three-cluster solution for test and replication data sets. Sensory profiles of the three clusters presented as mean z scores \pm 95% confidence interval for the test dataset ($n = 902$, A) and the validation dataset ($n = 233$, B). Note that z-transformation eliminates differences due to test site, gender and age. Positive z-scores indicate positive sensory signs (hyperalgesia), negative z-values indicate negative sensory signs (hypoesthesia, hypoalgesia). Dashed lines: 95% confidence interval for healthy participants ($-1.96 < z < +1.96$). Inserts show numeric pain ratings for DMA on a logarithmic scale (0-100) and frequency of PHS (0-3). Blue symbols: cluster 1 "sensory loss" (42% in A, 53% in B). Red symbols: cluster 2 "thermal hyperalgesia" (33% in A and B). Yellow symbols: cluster 3 "mechanical hyperalgesia" (24% in A, 14% in B). From (Baron et al., 2017).

Figure 4 illustrates the distinction of the three clusters in a 2D-scatter-plot and histograms of those two QST parameters that exhibited the best separation of

clusters: warm detection threshold (WDT) and mechanical pain sensitivity (MPS). Patients in cluster 1 had loss of pinprick sensitivity, while those in cluster 3 had pinprick hyperalgesia. Most patients in cluster 2 had WDT within the normal range of ± 1.96 z-values, while many of clusters 1 and 3 had hypoesthesia to warmth (z-values below -1.96). Although the k-means cluster separation was calculated in 13-dimensional space, this 2D-projection illustrates some of the main characteristics how the three clusters differ between each other. Partial overlap between clusters may also be due to two mechanisms present in the same patient. WDT and MPS were therefore chosen for simplified phenotyping subsequently.

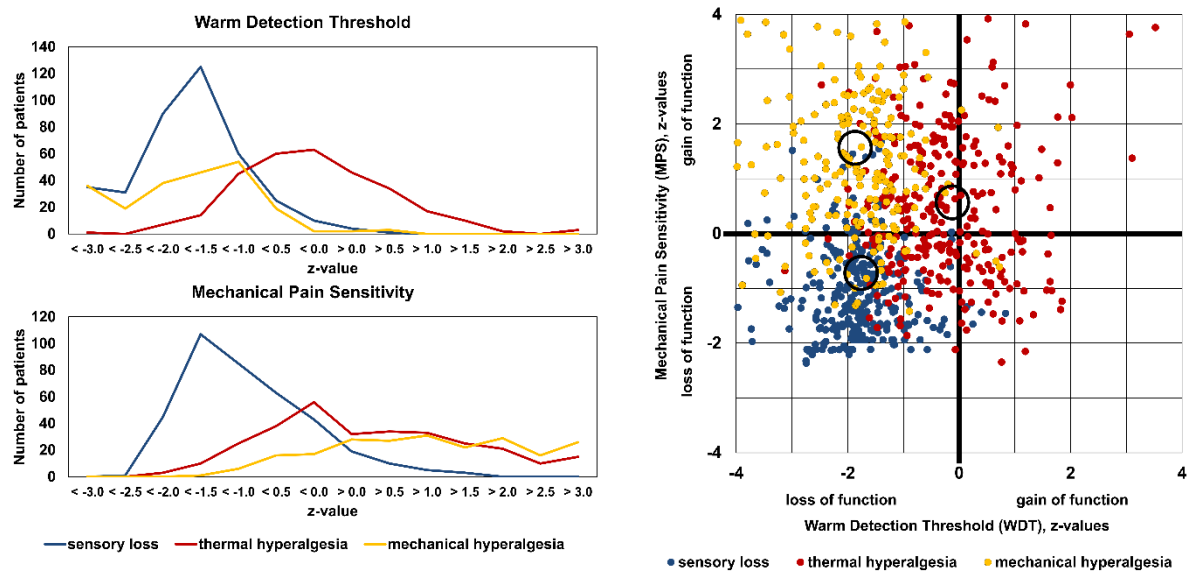


Figure 4: Cluster separation projected onto two-dimensional space. Histograms and scatter plot of the two QST-parameters that gave the best cluster separation: warm detection threshold (WDT) and mechanical pain sensitivity (MPS). Blue: cluster 1 "sensory loss" ($n=381$ patients), red: cluster 2 "thermal hyperalgesia" ($n=302$ patients), yellow: cluster 3 "mechanical hyperalgesia" ($n=219$ patients). Circles indicate centroids of each cluster. Modified from (Baron et al., 2017).

3.3.2 Patient characteristics of the three clusters

The patients' gender and mean age as well as pain intensity did not differ between the three groups (Table 4). Depressive symptoms occurred significantly more frequently in the "sensory loss" cluster. Spontaneous pain described by the patients as "stabbing" was comparable across the clusters but "burning" pain was significantly less frequent in the "thermal hyperalgesia" cluster.

	sensory loss	thermal hyperalgesia	mechanical hyperalgesia
n (%)	381 (42%)	302 (33%)	219 (24%)
age	59 ± 14	56 ± 14	59 ± 15
female	169 (39%)	152 (35%)	108 (25%)
depression	104 (47%) *	69 (31%)	49 (22%)
pain intensity	6.1 ± 3.1	5.8 ± 3.2	6.1 ± 3.0
burning pain	4.5 ± 3.4	4.3 ± 3.3	5.1 ± 3.2 *
stabbing pain	4.7 ± 3.2	4.3 ± 3.2	5.0 ± 3.0
sensory profile			
sensory loss	touch, thermal, pain	none	mostly thermal
hyperalgesia	none	mostly cold and heat	mostly pressure and pin
DMA	little	little	much
PHS	much	little	little
mechanisms			
sensory loss	small and large fibers	-	mostly small fibers
hyperalgesia	-	mostly peripheral sensitization	mostly central sensitization
ongoing pain	ectopic activity in damaged nociceptors or CNS	spontaneous activity in surviving nociceptors	(ectopic?) activity in nociceptors

Table 4: Cluster characteristics. * $p < 0.05$. From (Baron et al., 2017).

According to the published DFNS reference data, each QST parameter in each patient can be individually rated as within or outside of the 95% confidence interval of variability in healthy age- and gender-matched subjects. This analysis is presented in Figure 5. Of patients in cluster 1 ("sensory loss"), more than 50% had significant non-nociceptive sensory loss on an individual basis. Paradoxical heat sensation occurred in 40% and sensory loss for pain sensitivity was also prevalent, although at less than 50%.

Patients of cluster 2, in contrast, exhibited hardly any sensory loss (except for touch in about 20% of patients), but significant proportions of patients with hyperalgesia to various stimuli. Cold and heat hyperalgesia were only significant for this cluster, but - probably at least partly due to the substantial variability of CPT and HPT in healthy participants - all percentages were clearly below 50%.

Patients of cluster 3 were characterized by combination of loss of detection of non-nociceptive stimuli and hyperalgesia to noxious stimuli. However, in contrast to cluster 1 the sensory loss was more pronounced for small fiber function, i.e. diminished temperature perception but relatively preserved tactile perception, and hyperalgesia was only present for mechanical stimuli. DMA was present in the vast majority of these patients.

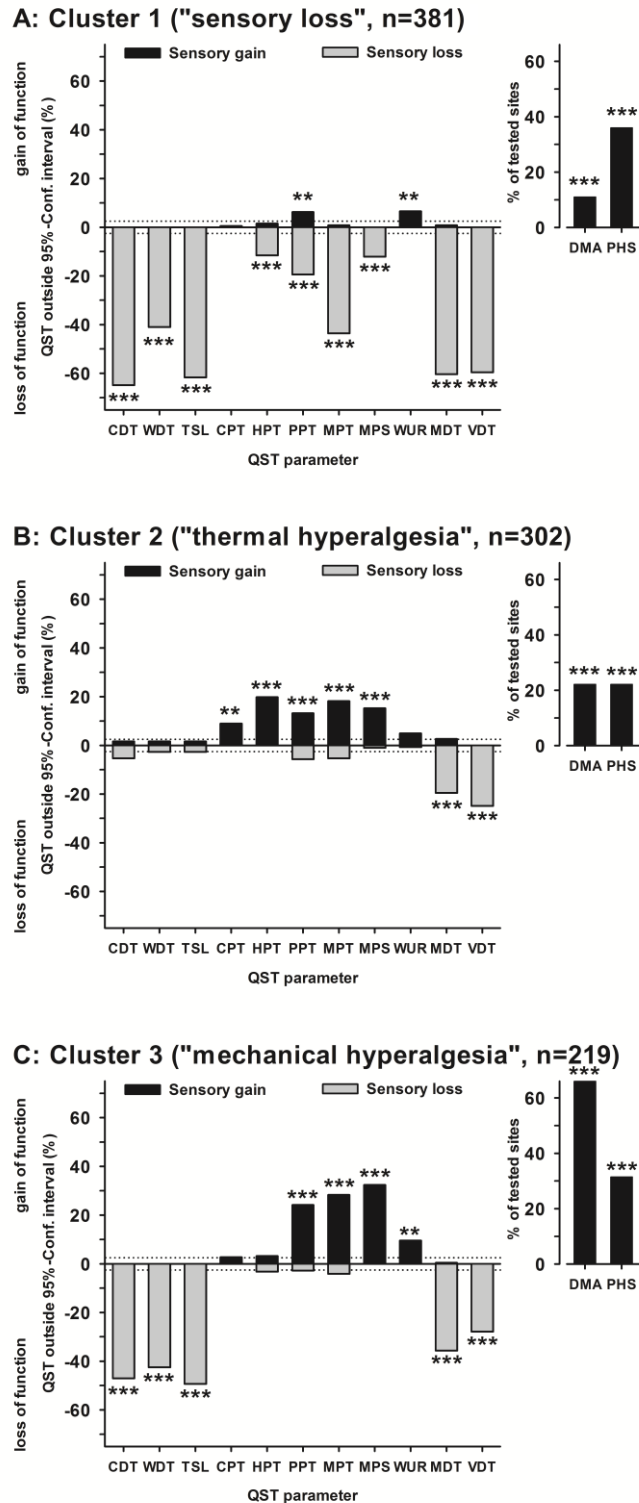


Figure 5: Frequencies of abnormal QST findings for the test dataset (n = 902). Each column gives the percentage of patients with abnormal findings for that particular QST parameter (outside the 95% confidence interval of healthy participants). Positive values indicate positive sensory signs (hyperalgesia), negative values indicate negative sensory signs (hypoesthesia, hypoalgesia). Dashed lines: Expected value for healthy participants ($\pm 2.5\%$). A: cluster 1 "sensory loss" (n=381 patients), B: cluster 2 "thermal hyperalgesia" (n=302 patients), C: cluster 3 "mechanical hyperalgesia" (n=219 patients). Significant compared to expected value (2.5%) on * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. From (Baron et al., 2017).

3.3.3 Distribution of clusters across etiologies

Figure 6 illustrates that in principle all three clusters were distributed across all four etiologies, demonstrating that the sensory signs of neuropathic pain that are produced by these etiologies overlap considerably. Each of the different etiologies, however, showed a characteristic pattern of sensory profiles: In PNI, patients with “thermal hyperalgesia” were significantly more frequent (40.1%) than patients with other sensory profiles. In PNP and RAD, “sensory loss” occurred in 51.8% and 42.7% of cases, respectively. Patients with PHN were concentrated in the “mechanical hyperalgesia” cluster (46.6%).

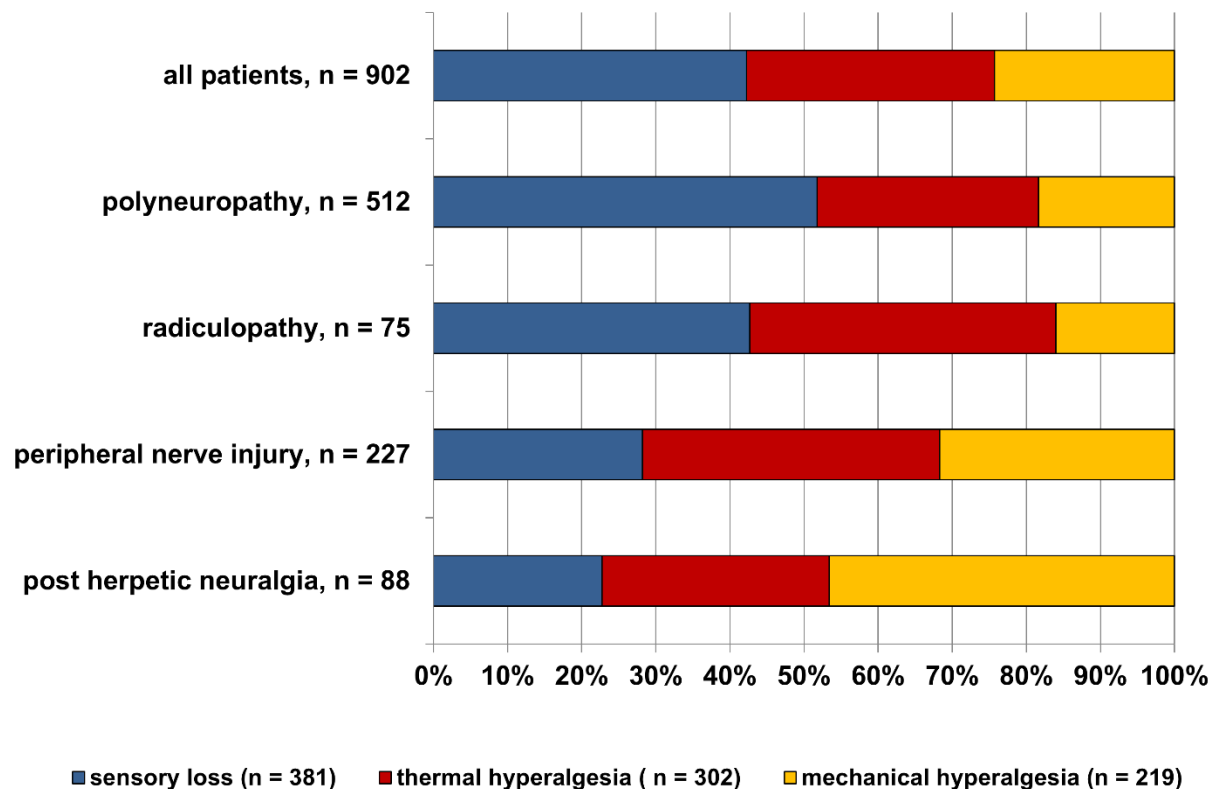


Figure 6: Distribution of the three clusters within each neuropathic pain etiology. Blue bars: cluster 1 "sensory loss" (n=381 patients), red bars: cluster 2 "thermal hyperalgesia" (n=302 patients), yellow bars: cluster 3 "mechanical hyperalgesia" (n=219 patients). Cluster 1 was most frequent in polyneuropathy, cluster 2 in peripheral nerve injury and radiculopathy, and cluster 3 in post herpetic neuralgia. From (Baron et al., 2017).

3.4 Individual sorting algorithm

[The following section has been taken in parts and modified from (Vollert et al., 2017a).]

3.4.1 Sorting algorithm

The sorting algorithm was based on the means and standard deviations for each phenotype identified in the cluster analysis (Table 5).

	sensory loss (μ (σ))	thermal hyperalgesia (μ (σ))	mechanical hyperalgesia (μ (σ))	healthy participants (μ (σ))
CDT	-2.42 (1.16)	-0.47 (1.04)	-2.03 (1.17)	0.00 (1.00)
WDT	-1.96 (0.96)	-0.25 (0.97)	-2.01 (1.14)	0.00 (1.00)
TSL	-2.23 (0.92)	-0.45 (0.93)	-2.10 (0.93)	0.00 (1.00)
CPT	-0.56 (0.81)	0.59 (1.09)	-0.15 (1.01)	0.00 (1.00)
HPT	-1.20 (0.87)	0.78 (1.45)	-0.67 (1.07)	0.00 (1.00)
PPT	-0.53 (1.56)	0.34 (1.56)	1.09 (2.02)	0.00 (1.00)
MPT	-1.60 (1.23)	0.42 (1.56)	0.86 (1.55)	0.00 (1.00)
MPS	-1.14 (0.81)	0.49 (1.35)	1.31 (1.41)	0.00 (1.00)
WUR	0.13 (1.04)	-0.01 (1.03)	0.21 (1.18)	0.00 (1.00)
MDT	-3.08 (4.94)	-0.91 (2.46)	-1.73 (2.48)	0.00 (1.00)
VDT	-2.88 (2.70)	-1.02 (1.84)	-1.18 (2.02)	0.00 (1.00)
PHS	0.72 (0.96)	0.63 (0.93)	0.44 (0.83)	0.00 (1.00)
DMA	0.24 (0.69)	1.67 (1.21)	0.54 (1.04)	0.00 (1.00)

Table 5: Mean QST z-values (μ) and standard deviations (σ , in brackets) for each of the 13 QST parameters separately for each of the three phenotypes. Values for healthy participants follow the definition of z-values: mean = 0 and standard deviation = 1. PHS is coded as pseudo-normally distributed with 0 = absence and 2 = presence, DMA is coded pseudo-normally distributed with 0 = absence, 2 = 0-1 (on a 0-100 numerical rating scale), and 3 = 1-100. From (Vollert et al., 2017a).

Individual allocation replicates the original cluster analysis (Baron et al., 2017) in 81% of the cases for the complete QST protocol using 13 parameters and in 76% of the cases using simplified phenotyping (only via WDT and MPS). Cohen's kappa

coefficient of agreement (scale: 0 = random classification, 1 = perfect agreement between methods) was 0.72 for the complete protocol and 0.63 for simplified phenotyping, both values may be categorized as “good”, although no universal guideline for interpreting Cohen’s kappa exists (Fleiss, 1973). Most common shifts were former sensory loss or thermal hyperalgesia to mechanical hyperalgesia (14% and 17%, respectively), and least common shift was former sensory loss to thermal hyperalgesia (<1%). Patient shift between phenotypes is shown in Table 6.

algorithm	cluster	sensory loss n = 381	thermal hyperalgesia n = 302	mechanical hyperalgesia n = 219
sensory loss n = 356 (356)		325 (301)	15 (29)	16 (26)
thermal hyperalgesia n = 267 (282)		3 (34)	235 (219)	29 (29)
mechanical hyperalgesia n = 279 (264)		53 (46)	52 (54)	174 (164)

Table 6: Crosstabulation of dominant phenotype identified using cluster analysis vs. the proposed new, individualized algorithm (rows) for full and simplified phenotyping (in brackets). Overall, both classifications revealed a strong concordance of solutions (81% of the cases, 76% for simplified phenotyping using only warm detection threshold and mechanical pain sensitivity). From (Vollert et al., 2017a).

3.4.2 Discrimination analysis against healthy participants

The ROC analysis for discrimination of patients against healthy participants is shown in Figure 7. The ROC-AUC value (scale: 0.5 – 1, 0.5: no discriminatory power, 1: perfect discrimination) for separating patients with neuropathic pain and healthy participants using the probability for being healthy was found to be 0.915 (95% CI: 0.898 – 0.932), indicating high discriminatory power. For simplified phenotyping, discriminatory power was significantly lower (0.785, 0.753 – 0.815). The Youden-Index was found to be highest at a probability of 64% - i.e., each subject with a probability value below 64% should be considered as a patient, and when above 64% as being healthy. For simplified phenotyping, the highest Youden-Index was found at a very similar value of 63% with similar sensitivity (74%) but very reduced

specificity (72%). Due to the high similarity of cut-offs, 64% was used for both full protocol and simplified phenotyping.

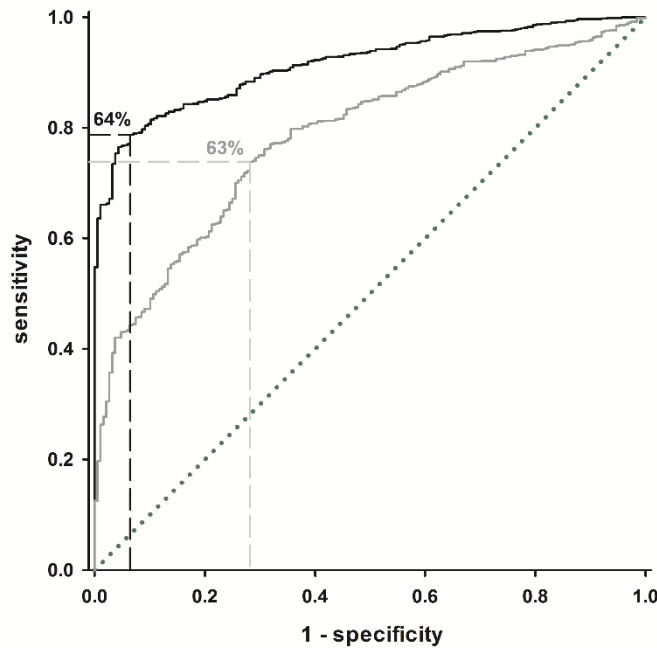


Figure 7: ROC (Receiver Operating Characteristics) analysis the discriminatory power of the healthy probability to separate between patients with neuropathic pain and healthy participants. Black line: full sensory testing, gray line: reduced protocol, using only warm detection threshold and mechanical pain sensitivity. The green dotted diagonal line indicates random classification (“coin flipping”). The area marked by dashed lines indicates the optimum ratio of sensitivity and specificity at 64% (reduced phenotyping: 63%) probability for being healthy. From (Vollert et al., 2017a).

Individual probabilities for each phenotype for patients and healthy participants are plotted in Figure 8. Clinically, abnormal QST values have been defined as outside 95% of the values found in healthy participants per parameter. As the QST protocol covers eleven normally distributed parameters, the chance of finding at least one abnormal parameter in the profile of a healthy participant is almost 50% (Vollert et al., 2015):

$$F(6): \quad p(\text{healthy}|\text{abnormal}) = \sum_{k=1}^{11} \binom{11}{k} 0.05^k 0.95^{11-k} \cong 43\%$$

Using the defined threshold, this probability could be reduced to 6% (94% of healthy participants detected correctly), sensitivity in detecting patients was 78% (i.e., 22% of

patients with neuropathic pain have a sensory profile with a probability for being healthy above 64%).

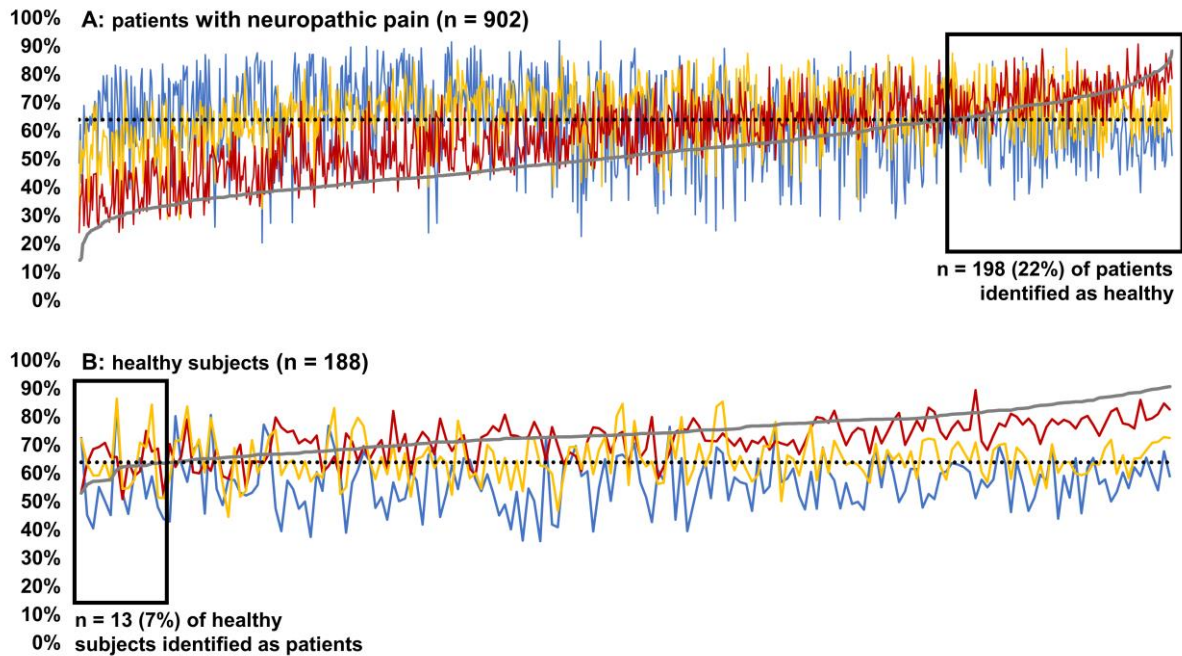


Figure 8: Phenotype probabilities and probability of being healthy for $n = 902$ patients with neuropathic pain and $n = 188$ healthy participants. Grey line: probability for being healthy, blue line: sensory loss, red line: thermal hyperalgesia, yellow line: mechanical hyperalgesia. Subjects on the x-axis are sorted by their individual probability of being healthy. Dotted line: a phenotype with a probability over 64% should be considered relevant in the individual patient. From (Vollert et al., 2017a).

3.4.3 Deterministic and probabilistic algorithm

To this point, we use a deterministic approach, i.e. each patient is allocated to exactly one phenotype. It is, however, possible that a patient may be allocated to more than one phenotype if each phenotype represents one set of mechanisms. So, with the cut-off determined for healthy participants above transferred onto patients, we can suggest two alternative versions of the algorithm, a deterministic one:

1. Calculate $F(5)$ for each of the 13 QST parameters. Use μ and σ from Table 5 for healthy.
2. Average the 13 probabilities. The resulting value is the probability for this patient to show a healthy profile.

3. Repeat steps 1 and 2, using μ and σ from Table 5 for sensory loss, thermal hyperalgesia, and mechanical hyperalgesia.
4. Allocate the patient to the phenotype with the **highest probability** value.

And a probabilistic version, where steps 1 – 3 remain identical and step four is exchanged with:

4. Sort the patient to **all phenotypes with a probability above 64%**. If the only probability over 64% is for being healthy or no phenotype reaches a probability of 64%, the patient should be excluded.

These two versions were used for all analyses below and are presented alongside. The simplified version of the algorithm is the same, except in step 1, only WDT and MPS are used instead of all 13 QST parameters, as these parameters have shown to explain the largest part of variability between the three phenotypes in our previous analysis (Baron et al., 2017).

3.4.4 Effectiveness for treatment with oxcarbazepine

Oxcarbazepine is a sodium channel blocker acting at the endings of small peripheral nerves, so it is assumed that it would be inferior as pain medication for patients suffering from deafferentation-related neuropathic pain. Therefore, based on the mechanistic assumptions from Table 4, oxcarbazepine should be more effective in the “thermal hyperalgesia” phenotype in comparison to the “sensory loss” and “mechanical hyperalgesia” phenotype, which both present significant loss of thermal detection, hinting towards loss of small fiber function. Within a cohort of $n = 83$ patients suffering from peripheral neuropathic pain whose pain relief after placebo and oxcarbazepine treatment is known (Demant et al., 2014), $n = 32$ patients presented the thermal hyperalgesia phenotype when the deterministic version of the algorithm was applied. In the probabilistic version, $n = 39$ patients showed a relevant probability over 64% for the thermal hyperalgesia phenotype. Overlap between deterministic and probabilistic “thermal hyperalgesia” phenotype and the “irritable nociceptor” phenotype used in the original study is shown in Table 7.

	deterministic thermal hyperalgesia	probabilistic thermal hyperalgesia
all patients (n = 83)	32	39
irritable nociceptor (n = 31)	24	22
non-irritable nociceptor (n = 52)	12	17

Table 7: Crosstabulation of deterministic and probabilistic identification of thermal hyperalgesia phenotype in relation to irritable and non-irritable nociceptor phenotype in (Demant et al., 2014).

Mean pain reduction in the verum phase in the subgroup of patients with “irritable nociceptor” phenotype of the original study (n = 31) was 1.4 (95% CI: 0.7 – 2.8, NRS 0 - 10). Mean pain reduction under oxcarbazepine was 1.0 (95% CI: 0.5 – 2.0; treatment-phenotype interaction: $f = 3.384$ ($p = 0.020$); $p = 0.437$ in comparison to the original “irritable nociceptor” group) for the deterministically phenotyped subgroup and 1.1 (95% CI: 0.6 – 2.3; treatment-phenotype interaction: $f = 3.431$ ($p = 0.066$); $p = 0.475$ in comparison to the original “irritable nociceptor” group) for the subgroup identified with the probabilistic algorithm.

The number-needed-to-treat was 3.9 (95% CI: 2.3 – 11.5) in the IN subgroup, 6.2 (95% CI: 3.3 – 122.1) in the subgroup identified by the deterministic algorithm, and 6.7 (95% CI: 3.7 – 24.6) in the subgroup identified by the probabilistic algorithm.

Overall, both versions of the algorithm provided a subgroup in which oxcarbazepine was only to a non-significant degree less effective than the initially used “irritable nociceptor” classification. The treatment-phenotype interaction variable in the mixed-effects model was only significant for the deterministic phenotyping, not the probabilistic phenotyping, but it should be stated that the variance found between the three subgroups is well within the margin of random sampling error, showing that cluster-based phenotyping is non-inferior to “irritable nociceptor” phenotyping.

3.5 Frequency of phenotypes in clinical entities

[The following section has been taken in parts and modified from (Vollert et al., 2017a).]

An overview of the frequency of each phenotype across the etiologies, full or simplified phenotyping and deterministic and probabilistic algorithm is presented in Table 8, details in the chapters below. Frequency of phenotypes and overlap between phenotypes for each clinical entity for the full protocol is displayed in Venn and bar diagrams in Figure 9.

Phenotyping protocol:	full		simplified	
	deterministic	probabilistic	deterministic	probabilistic
Diabetic polyneuropathy (n = 151)				
<i>Healthy profile</i>	4%	14%	4%	12%
<i>Sensory loss</i>	64%	82%	61%	64%
<i>Thermal hyperalgesia</i>	13%	33%	9%	20%
<i>Mechanical hyperalgesia</i>	19%	75%	26%	36%
Peripheral nerve injury (n = 335)				
<i>Healthy profile</i>	8%	19%	14%	22%
<i>Sensory loss</i>	24%	29%	26%	22%
<i>Thermal hyperalgesia</i>	31%	44%	24%	33%
<i>Mechanical hyperalgesia</i>	37%	52%	36%	36%
Post-herpetic neuralgia (n = 97)				
<i>Healthy profile</i>	6%	19%	9%	20%
<i>Sensory loss</i>	20%	22%	23%	12%
<i>Thermal hyperalgesia</i>	30%	39%	28%	35%
<i>Mechanical hyperalgesia</i>	44%	49%	40%	34%

Table 8: Frequency of each phenotype in diabetic polyneuropathy, peripheral nerve injury and post-herpetic neuralgia, separately for the deterministic and probabilistic algorithm, and for full and simplified phenotyping. From (Vollert et al., 2017a).

3.5.1 *Deterministic*

Most common phenotype in diabetic polyneuropathy was sensory loss (n = 96, 64%), while in 20 (13%) patients thermal hyperalgesia and in 29 (19%) patients mechanical hyperalgesia was the dominant phenotype. Six (4%) patients presented a profile most similar to healthy participants.

In peripheral nerve injury, mechanical and thermal hyperalgesia were frequent (n = 125, 37% and n = 105, 31%, respectively), while sensory loss was less prominent (n = 82, 24%). Twenty-three (7%) patients showed highest probability for a healthy QST profile. Post herpetic neuralgia was mostly characterized by the mechanical hyperalgesia phenotype (n = 43, 44%), followed by thermal hyperalgesia (n = 29, 30%) and sensory loss (n = 19, 20%). Six (6%) patients had the highest probability for a QST profile similar to healthy participants.

3.5.2 *Probabilistic*

Of the diabetic polyneuropathy cohort, 4 patients (3%) were not sorted to any phenotype nor healthy and had to be excluded. Twenty-one (14%) patients showed a relevant probability for the healthy profile, all of them were additionally assigned to a phenotype. Most common phenotype (n = 124, 82%) was sensory loss, while in 50 (33%) patients thermal hyperalgesia and in 113 (75%) patients mechanical hyperalgesia was a prominent phenotype, the latter in significant difference to the deterministic algorithm (19%). Twenty-seven patients (18%) were possible to assign to all three phenotypes and 86 (57%) to two phenotypes, with the largest overlap between sensory loss and mechanical hyperalgesia.

In peripheral nerve injury, 70 (21%) patients were not assigned to any phenotype and 2 (<1%) only to the healthy profile, these patients were all excluded. Sixty-three (19%) patients were assigned to the healthy profile and at least one additional phenotype. Overall, mechanical hyperalgesia was most frequent, followed by thermal hyperalgesia (n = 173 (52%) and n = 146 (44%), respectively), while sensory loss was less prominent (n = 98, 29%). 17 patients (5%) were allocated to all three phenotypes, 120 (36%) to two phenotypes, with sensory loss and thermal hyperalgesia showing the least overlap.

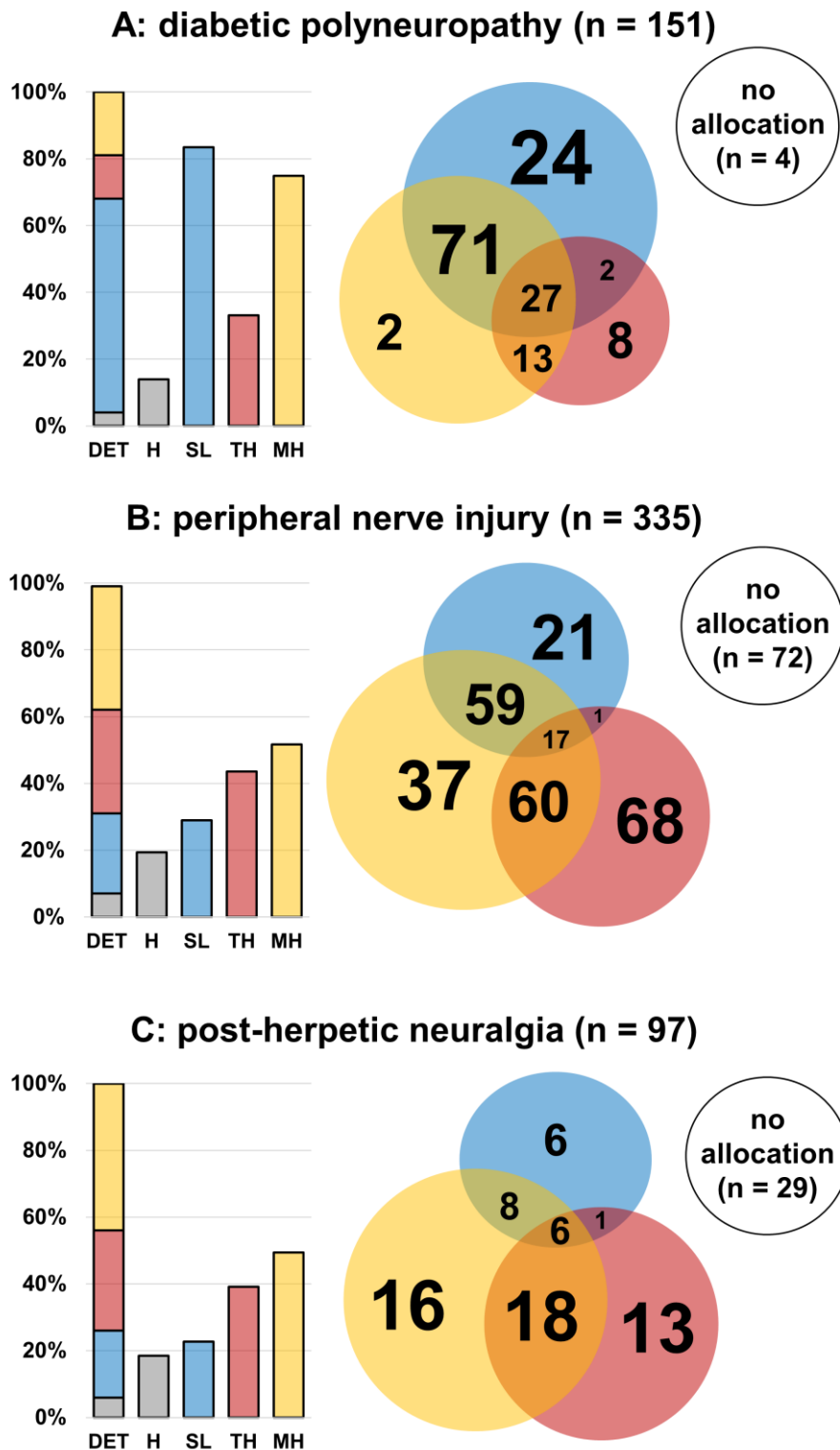


Figure 9: Venn and bar diagrams of phenotype frequency and overlap between phenotypes for A) diabetic polyneuropathy, B) peripheral nerve injury and C) post-herpetic neuralgia. Bars are to scale, size of the circles and overlaps is illustrative, not to scale. Grey: healthy, blue: sensory loss, red: thermal hyperalgesia, yellow: mechanical hyperalgesia. First bar (DET): deterministic algorithm (adds to 100%), three subsequent bars: probabilistic approach (a patient may be allocated to more than one phenotype, percentages are not additive). H: healthy, SL: sensory loss, TH: thermal hyperalgesia, MH: mechanical hyperalgesia. From (Vollert et al., 2017a).

In post-herpetic neuralgia, 29 (30%) patients were not possible to assign to any phenotype, these patients were excluded. Eighteen patients showed a relevant probability for a healthy profile, all of them were additionally sorted to at least one phenotype. Overall, patients with post-herpetic neuralgia were mostly characterized by the mechanical hyperalgesia phenotype (n = 48, 49%), followed by thermal hyperalgesia (n = 38, 39%) and sensory loss (n = 21, 22%). Six patients (6%) were possible to assign to all three phenotypes, 27 (28%) to two phenotypes, with the largest overlap between thermal and mechanical hyperalgesia.

3.5.3 Simplified phenotyping

Overall, 57% of patients with diabetic polyneuropathy, 62% of patients with peripheral nerve injury and 58% of patients with post-herpetic neuralgia were sorted into the same phenotype allocated when the full protocol was applied. The sensitivity of the simplified algorithm, however, is dependent on a combination of phenotype of interest and the clinical entity under study: in diabetic polyneuropathy, 74% of sensory loss patients were correctly identified, but only 48% of patients with thermal hyperalgesia and even less 43% of patients with mechanical hyperalgesia. In patients with peripheral nerve injury, allocation accuracy was more balanced between phenotypes (75% for sensory loss, 60% for thermal hyperalgesia, 64% for mechanical hyperalgesia). In patients with post-herpetic neuralgia, sensitivity was very low for sensory loss (24%), and better for thermal (76%) and mechanical hyperalgesia (56%).

3.6 Sample size recommendations

[The following section has been taken in parts and modified from (Vollert et al., 2017a).]

Estimated sample sizes for parallel or crossover design, the three phenotypes and the three etiologies of neuropathic pain are presented in Table 9 for both full and simplified phenotyping, applying the deterministic or probabilistic version of the algorithm.

In summary, for parallel study design, either the estimated effect size of the treatment needs to be high (>0.7) or only phenotypes that are frequent in the clinical entity

under study can realistically be performed. For crossover design, populations under 200 patients need to be screened for all phenotypes and clinical entities with a minimum estimated treatment effect size 0.5.

Low sensitivity of the simplified algorithm is linked to low frequency of certain phenotypes (esp. thermal hyperalgesia in diabetic polyneuropathy and sensory loss in post-herpetic neuralgia), leading to higher numbers of patients that need to be screened.

Study design	parallel			crossover		
	Effect size	0.3	0.5	0.7	0.3	0.5
Diabetic polyneuropathy						
<i>Sensory loss</i>	550 (577)	200 (210)	106 (111)	141 (148)	53 (56)	30 (31)
	429 (550)	156 (200)	83 (106)	110 (141)	41 (53)	23 (30)
<i>Thermal hyperalgesia</i>	2708 (3911)	985 (1422)	523 (756)	692 (1000)	262 (378)	146 (211)
	1067 (1760)	388 (640)	206 (340)	273 (450)	103 (170)	58 (95)
<i>Mechanical hyperalgesia</i>	1853 (1354)	674 (492)	358 (262)	474 (346)	179 (131)	100 (73)
	469 (978)	171 (356)	91 (189)	120 (250)	45 (94)	25 (53)
Peripheral nerve injury						
<i>Sensory loss</i>	1467 (1354)	533 (492)	283 (262)	375 (346)	142 (131)	79 (73)
	1214 (1600)	441 (582)	234 (309)	310 (409)	117 (155)	66 (86)
<i>Thermal hyperalgesia</i>	1135 (1467)	413 (533)	219 (283)	290 (375)	110 (142)	61 (79)
	800 (1067)	291 (388)	155 (206)	205 (273)	77 (103)	43 (58)
<i>Mechanical hyperalgesia</i>	951 (978)	346 (356)	184 (189)	243 (250)	92 (94)	51 (53)
	677 (978)	246 (356)	131 (189)	173 (250)	65 (94)	37 (53)
Post-herpetic neuralgia						
<i>Sensory loss</i>	1760 (1530)	640 (557)	340 (296)	450 (391)	170 (148)	95 (83)
	1600 (2933)	582 (1067)	309 (567)	409 (750)	155 (283)	86 (158)
<i>Thermal hyperalgesia</i>	1173 (1257)	427 (457)	227 (243)	300 (321)	113 (121)	63 (68)
	903 (1006)	328 (366)	174 (194)	231 (257)	87 (97)	49 (54)
<i>Mechanical hyperalgesia</i>	800 (880)	291 (320)	155 (170)	205 (225)	77 (85)	43 (48)
	718 (1035)	261 (376)	139 (200)	184 (265)	69 (100)	39 (56)

Table 9: Number of patients, that need to be screened to find a sub-population with a given phenotype large enough to conduct a study with a power of 80% with an alpha-level of 0.05 and a given effect size. First row: deterministic algorithm, second row: probabilistic algorithm. Values in brackets show the number needed to be screened with the simplified protocol. Numbers in bold indicate that 200 or less patients need to be screened. From (Vollert et al., 2017a).

3.7 Subgrouping human surrogate models

[The following section has been taken in parts and modified from (Vollert et al., 2017b).]

The analysis comprised a total of $n=657$ healthy subjects that participated in studies on human surrogate models: nine distinct models (two of them in the area of primary and secondary hyperalgesia) at five centers (Table 1). About 44% of subjects were female (291/657); gender distribution was not homogenous across models, as for the UVB, cutaneous HFS and menthol models predominantly males were recruited. Age ranges were mostly lower than in neuropathic pain populations. These factors were accounted for by normalizing all QST data to gender-specific and age-specific reference data.

Fig. 10 shows z-profiles of the six models used for defining phenotype means. As human surrogate models of nerve blocks, both nerve compression and topical lidocaine led to substantial loss in thermal and mechanical detection thresholds (Fig. 10A). For CDT and MDT in the A-fiber block, this loss was almost complete, reaching a mean z-value beyond -5, i.e. beyond five standard deviations of normal detection. A-fiber block was also associated with pinprick hyperalgesia and paradoxical heat sensations.

As human surrogate models of primary hyperalgesia, topical capsaicin and UVB sunburn both induced substantial heat and mechanical hyperalgesia (Fig. 10B). Heat hyperalgesia was more pronounced for capsaicin, mechanical hyperalgesia more pronounced for UVB. Capsaicin also induced loss of cold detection and cold pain.

As human surrogate models of secondary hyperalgesia, cutaneous HFS and i.d. capsaicin injection led to mechanical hyperalgesia and thermal sensory deficits (Fig. 10C). Mechanical hyperalgesia and DMA were more pronounced in the capsaicin injection model.

Table 10 compares mean z-scores normalized to published reference data of healthy subjects with effect sizes in intra-individual comparison to untreated control areas. Intra-individual comparisons mostly confirmed the patterns of negative or positive sensory signs of z-values, but the loss of thermal detection for i.d. capsaicin and cutaneous HFS (Fig. 10C) appears to be overestimated in z-values due to the non-standard test area.

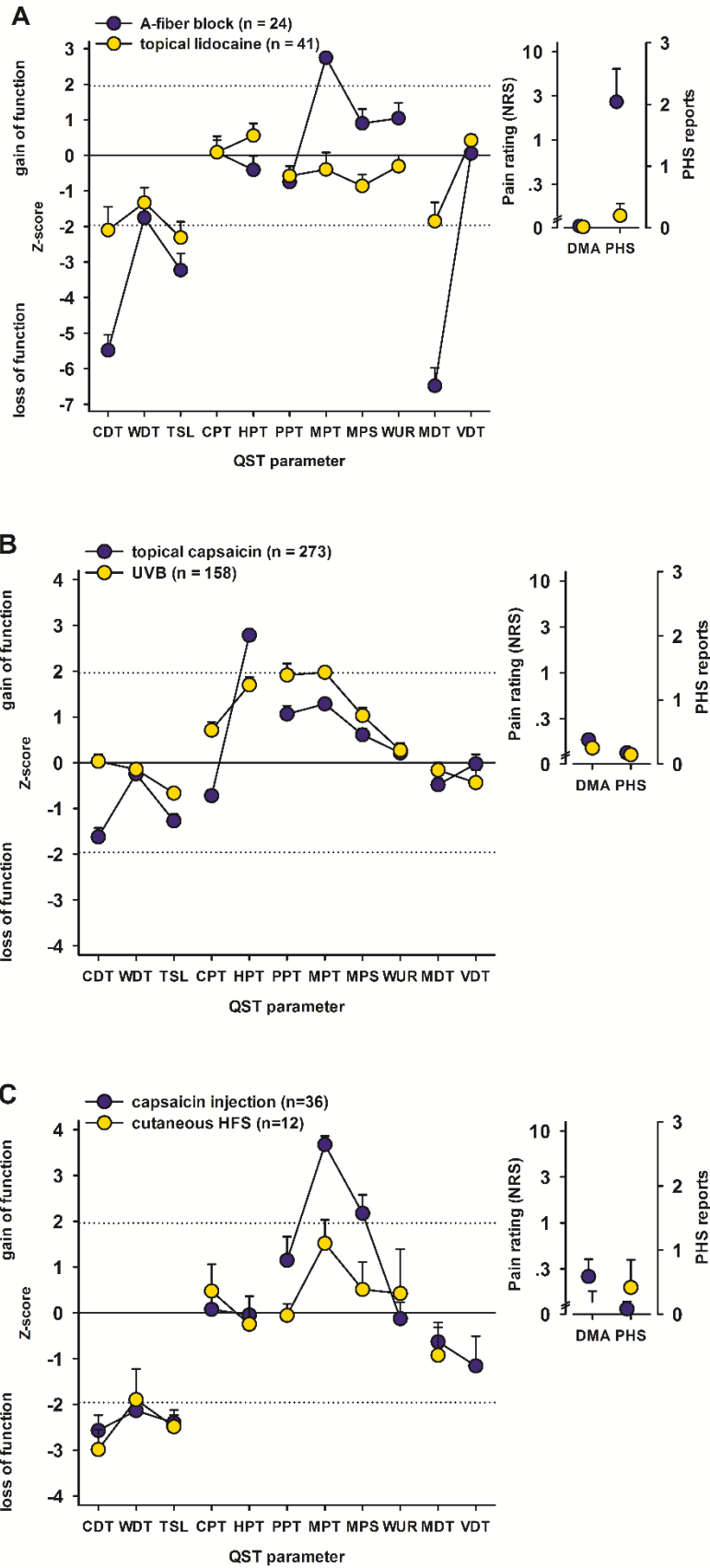


Figure 10: Deep sensory profiling of human surrogate models of a priori mechanisms. A: nerve blocks, B: primary hyperalgesia, C: secondary hyperalgesia. From (Vollert et al., 2017b).

		CDT	WDT	TSL	CPT	HPT	PPT	MPT	MPS	WUR	MDT	VDT
A-fiber block	z	-5.49	-1.75	-3.23	0.08	-0.41	-0.75	2.74	0.90	1.28	-6.48	0.06
	D	-4.36	-0.29	-1.90	0.69	0.28	-0.46	2.68	1.07	0.99	-4.83	0.02
Topical lidocaine	z	-2.10	-1.33	-2.31	0.09	0.56	-0.58	-0.40	-0.86	-0.37	-1.85	0.42
	D	-1.40	-1.21	-1.69	-0.59	0.05	-0.36	-1.27	-0.99	-0.36	-0.94	0.01
Topical capsaicin	z	-1.62	-0.24	-1.27	-0.72	2.79	1.06	1.29	0.61	0.21	-0.48	-0.02
	D	-1.22	-0.32	-0.84	-0.75	2.57	0.28	0.96	0.69	0.10	-0.30	0.08
UVB	z	0.03	-0.14	-0.67	0.71	1.70	1.92	1.97	1.03	0.27	-0.16	-0.42
	D	1.62	-0.23	0.40	0.18	1.30	0.78	2.80	1.39	-0.96	0.63	0.42
capsaicin i.d.	z	-2.57	-2.14	-2.40	0.08	-0.05	1.15	3.67	2.18	-0.13	-0.63	-1.16
	D	-0.49	-0.22	-0.29	-0.06	0.15	0.69	2.50	1.66	-0.20	0.28	0.04
cutaneous HFS	z	-2.98	-1.89	-2.49	0.48	-0.25	-0.06	1.52	0.51	0.42	-0.93	
	D	-0.69	-0.17	-0.40	-0.25	-0.18	0.49	0.91	0.90	-0.05	0.81	
2° hyperalgesia topical capsaicin	z	-1.64	-1.30	-1.48	0.48	0.65	1.07	1.71	1.10	0.14	-1.51	-2.18
	D	-0.40	-0.44	-0.34	0.06	0.31	0.22	0.73	1.04	-0.19	0.16	0.12
2° hyperalgesia UVB	z	-0.97	0.05	-0.74	0.59	0.58	1.10	0.56	0.14	1.24	-0.89	-1.11
	D	0.43	-0.02	0.30	0.07	0.33	0.10	0.73	0.41	0.05	-0.33	0.03
muscular HFS	z	-0.86	0.12	-0.41	0.75	0.83	1.01	0.52	0.08	0.49	-0.99	-0.86
	D	0.06	-0.04	-0.07	0.21	0.07	-0.13	-0.34	-0.15	0.04	0.04	0.00
topical menthol	z	-1.00	-0.72	-0.08	1.08	0.03	-0.70	1.33	0.25	-0.28	0.34	0.29
	D	-0.63	-0.41	0.07	1.94	0.67	0.32	2.59	0.80	-0.27	0.29	0.15
topical lidocaine + capsaicin	z	-3.31	-1.74	-3.09	-1.23	2.08	0.26	-0.07	-0.65	-0.20	-1.30	0.49
	D	-2.18	-1.57	-2.49	-2.49	2.07	0.21	-0.67	-0.30	-0.14	-0.53	-0.24

Table 10: z-scores and effect sizes (D) of human surrogate models. z-scores: boundaries 1.0 and 1.96. D values boundaries above 0.5 and 0.8. From (Vollert et al., 2017b).

The QST profiles of the remaining models are presented in Figure 11. Fig 11A shows the QST profiles of surrounding skin from two models where secondary hyperalgesia is either controversial (UVB) or known to be mild (topical capsaicin). While the area of secondary hyperalgesia of topical capsaicin displays sensory loss in the z-profile, these effects are much smaller in the intra-individual comparisons with untreated skin, again suggesting an overestimation in z-values due to the non-standard test area (Table 10). Topical menthol has been introduced to induce cold hyperalgesia, and muscle HFS to induce deep hyperalgesia (Fig. 11B); sensory changes in both models were mild. Fig. 11C shows a combined model of nerve block and primary hyperalgesia.

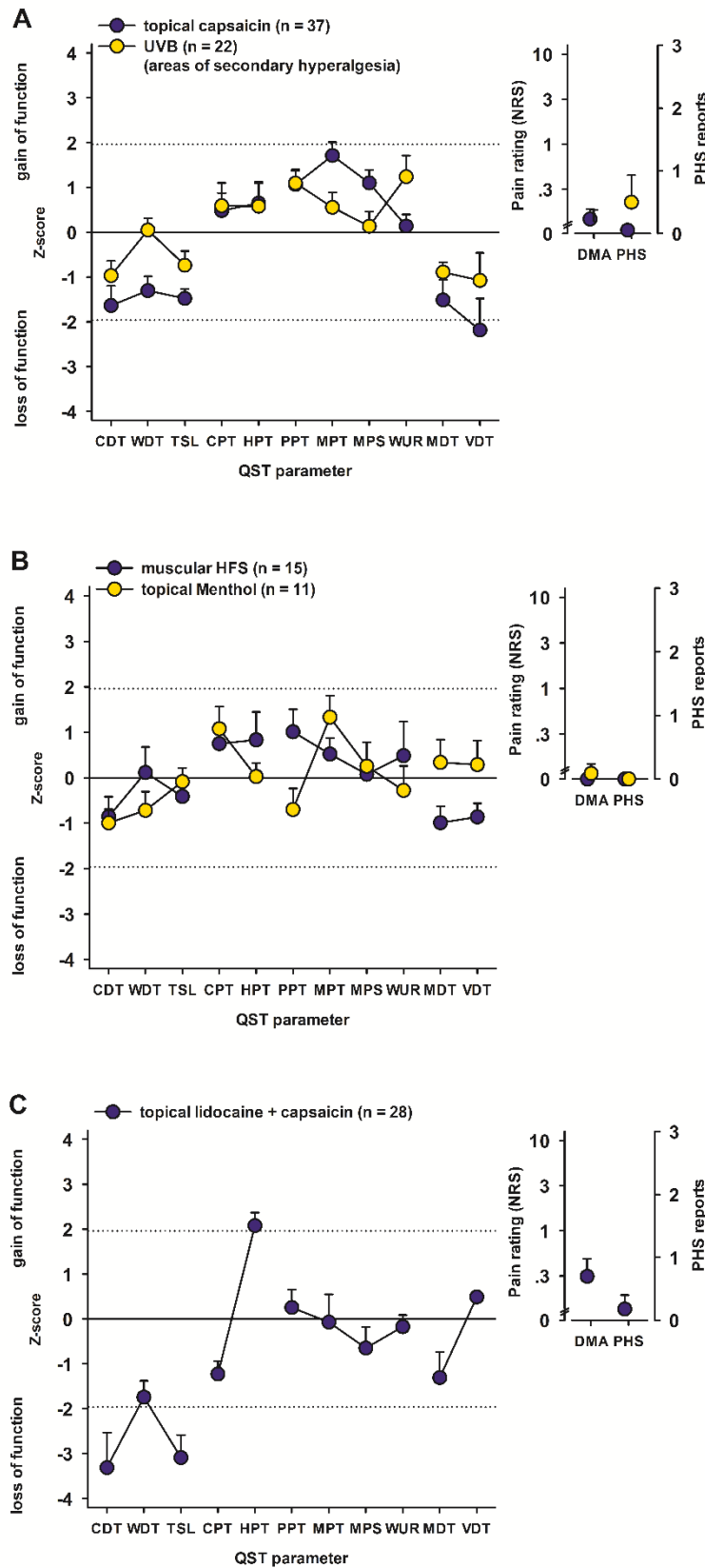


Figure 11: Deep sensory profiling of additional human surrogate models. A: secondary hyperalgesia surrounding topical capsaicin or UVB treated areas, B: topical menthol (cold hyperalgesia model) and muscle HFS (deep pain model), C: sequentially combined topical application of lidocaine and capsaicin. From (Vollert et al., 2017b).

3.7.1 *Cluster analysis of human surrogate models*

Unlike for patients suffering from neuropathic pain, the cluster analysis of human surrogate models did not reveal meaningful results. All solutions for 2 - 9 clusters presented with a high count of negative silhouettes (above 10%, see Table 11), which indicates solutions that are mere statistical artefacts rather than separate clusters and had to be excluded according to the quality criteria for acceptable solutions in our protocol. The two clusters solution scored highest in mean silhouette width and ARI in comparison to hierarchical clustering, but also showed a frequency of negative silhouettes above the cut-off of 10%. The ten clusters solution scored highest in the remaining quality criteria, still, ten is a suspiciously high number considering that only eleven models are included.

In addition, all solutions from 4 - 10 clusters were unstable in the k-means clustering, i.e. the algorithm presented various solutions in various runs instead of coming to the same solution for a given number of clusters, indicating various local optimum solutions rather than one globally optimal solution. Indices for comparison to methods of distinct mathematical background (hierarchical and Bayesian clustering) presented a similarity between methods far lower than it was the case for clustering of patients.

Therefore, for human surrogate models, no subgrouping based on a hypothesis-free cluster analysis of sensory profiles can be applied.

n (cluster)	silhouette width:			comparison to hierarchical:			comparison to EM:	
	mean °	minimum mean per cluster°°	negative (%)°°°	ARI*	AVI**	ARI*	AVI**	BIC***
2	0.29	0.06	14.5%	0.17	0.52	0.28	0.41	0
3	0.26	0.10	12.2%	0.17	0.47	0.23	0.45	747.25
4	0.21	0.13	12.5%					
5	0.20	0.12	11.3%					
6	0.18	0.10	11.9%					
7	0.17	0.09	16.1%					
8	0.20	0.11	14.8%					
9	0.21	0.14	11.6%					
10	0.21	0.11	9.7%	0.20	0.21	0.23	0.22	1338.85

Table 11: Decision on the number of clusters for human surrogate models. °Mean silhouette width per cluster. A value below zero indicates clusters that do not separate from other clusters. Measure of discriminatory power (0 to 1). 0: no discrimination, 1: perfectly separated clusters (high values are preferred). °°Measure of fragmentation of solution (-1 to +1). -1: cluster that is solely a fragment, +1: a solution that is not fragmented (solutions with values below zero were discarded) °°°Measure of fragmentation of solution (0% to 100%). 0%: no fragmentation, 100% a completely fragmented solution (solutions with values above 10% were discarded) *ARI (Adjusted Rand Index): Measure of similarity (0 to 1). 0: only random identity, 1: perfect identity (high values are preferred) **AVI (Adjusted Variation of Information): Measure of dissimilarity (0 to 1). 0: no dissimilarity, 1: strong dissimilarity (low values are preferred) ***Delta-BIC (Bayesian Information Criterion): Measure of gain of information by increasing cluster number. If delta-BIC > 10, the higher cluster number is recommended.

3.7.2 Pattern-based sorting algorithm

By random number assignment, 49% of subjects from the six surrogate models with clearly defined mechanism (topical lidocaine, A-fiber block, topical capsaicin, UVB radiation, capsaicin injection and cutaneous HFS) were assigned to the training data set (n = 265), which defined mean values and standard deviations for the sorting algorithm (Table 12).

	nerve block (μ (σ))	primary hyperalgesia (μ (σ))	secondary hyperalgesia (μ (σ))	healthy participants (μ (σ))
CDT	-2.94 (2.36)	-1.05 (1.59)	-2.71 (1.07)	0.00 (1.00)
WDT	-1.36 (1.27)	-0.23 (1.04)	-1.87 (0.89)	0.00 (1.00)
TSL	-2.55 (1.52)	-1.03 (1.12)	-2.28 (0.84)	0.00 (1.00)
CPT	0.06 (1.13)	-0.12 (1.31)	0.33 (1.16)	0.00 (1.00)
HPT	0.52 (1.23)	2.37 (1.14)	-0.14 (1.24)	0.00 (1.00)
PPT	-0.63 (0.97)	1.35 (1.66)	0.63 (1.45)	0.00 (1.00)
MPT	0.86 (1.87)	1.54 (1.08)	3.12 (0.97)	0.00 (1.00)
MPS	-0.15 (1.38)	0.71 (1.21)	1.95 (1.19)	0.00 (1.00)
WUR	0.25 (1.29)	0.18 (1.01)	0.31 (1.15)	0.00 (1.00)
MDT	-3.24 (2.89)	-0.32 (1.13)	-0.80 (1.38)	0.00 (1.00)
VDT	0.22 (1.06)	-0.23 (1.93)	-1.08 (1.83)	0.00 (1.00)
PHS	0.06 (0.34)	0.81 (1.07)	1.12 (1.11)	0.00 (1.00)
DMA	0.48 (0.86)	0.13 (0.49)	0.16 (0.54)	0.00 (1.00)

Table 12: Mean QST z-values (μ) and standard deviations (σ , in brackets) for each of the 13 QST parameters separately for each of the three mechanistic subgroups. Values for healthy participants follow the definition of z-values: mean = 0 and standard deviation = 1. PHS is coded as pseudo-normally distributed with 0 = absence and 2 = presence, DMA is coded pseudo-normally distributed with 0 = absence, 2 = 0-1 (on a 0-100 numerical rating scale), and 3 = 1-100. From (Vollert et al., 2017b).

Individual allocation by the deterministic sorting algorithm replicated the a priori assignment of surrogate models in 79% of the cases for training set and in 81% of the cases for the test set (remaining 279 subjects from the same surrogate models). Cohen's kappa coefficient of agreement (scale: 0 = random classification, 1 = perfect agreement between methods) was 0.54 (95% confidence interval: 0 – 1) for the training set and 0.56 (95% confidence interval: 0 – 1) for the test set, both values may be categorized as “good”, although no universal guideline for interpreting Cohen's kappa exists (Fleiss, 1973), and the confidence intervals reaching from zero to one indicate that the profiles are so unevenly distributed that these kappa values should not be overinterpreted. Most common shifts were primary or secondary hyperalgesia to nerve block (18% and 27%, respectively), and least common shifts

were nerve block to secondary hyperalgesia and secondary to primary hyperalgesia (both <1%). Shifts between original and algorithmic assignment is shown in Table 13.

algorithm	original	nerve block	primary hyperalgesia	secondary hyperalgesia
nerve block		29 / 30	4 / 2	0 / 0
primary hyperalgesia		35 / 42	162 / 178	10 / 4
secondary hyperalgesia		8 / 5	0 / 1	17 / 17

Table 13: Crosstabulation of dominant phenotype identified using the algorithm (columns) vs. the original assignment (rows) for training and test dataset (in brackets). Overall, concordance of the algorithm to original assignment was strong (79% of the cases for the training and 81% for the test dataset, note that the even higher concordance for the test dataset compared to training data indicates a very robust effect). From (Vollert et al., 2017b).

3.7.3 Deterministic and probabilistic sorting

The z-profiles of each sensory subgroup are shown in Figure 12. Forty-one subjects showed a sensory profile, that was most similar to untreated healthy skin, although part of a surrogate model (Fig. 12A). The subjects sorted to the nerve block profile were mostly characterized by loss of cold and mechanical detection (CDT and MDT, both representing A-fiber function, Fig. 12A). Loss of vibration detection was not detectable, which is due to a limitation in the A-fiber block, which only affects a small area, while vibration can be sensed by rapidly adapting mechanoreceptors situated beyond this area. Primary hyperalgesia (Fig. 12B) was characterized mostly by heat hyperalgesia, while secondary hyperalgesia presented mostly mechanical hyperalgesia, and also loss of thermal detection.

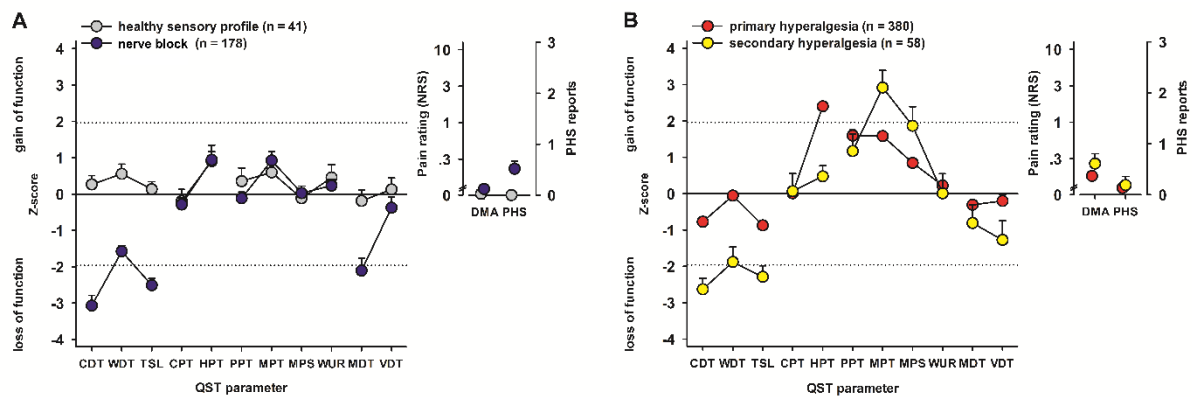


Figure 12: Average profiles per mechanism-stratified human surrogate models. Using the deterministic algorithm, 657 healthy subjects that had undergone human surrogate models of chronic pain were sorted into one of four categories of sensory profiles. A: nerve blocks vs. healthy profiles, B: primary vs. secondary hyperalgesia profiles. Positive z-scores indicate positive sensory signs (hyperalgesia), negative z-values indicate negative sensory signs (hypoesthesia, hypoalgesia). Dashed lines: 95% confidence interval for healthy participants ($-1.96 < z < +1.96$). Inserts show numeric pain ratings for DMA on a logarithmic scale (0-100) and frequency of PHS (0-3). From (Vollert et al., 2017b).

In the probabilistic sorting algorithm, the percentage of sensory profiles compatible being from normal skin increased to 187 (28%) vs. 41 (6%) in the deterministic version of the algorithm. The highest frequency of healthy profiles was found in muscular HFS and topical menthol (see Figure 13 for all frequencies per model and prototypic mechanistic profile), which also had the mildest sensory changes in their averaged QST profiles (Fig. 11B).

QST profiles compatible with the nerve block profile increased to 352 cases (54%) in the probabilistic algorithm vs. 178 (27%) in the deterministic version. Apart from the defining models of A-fiber block and topical lidocaine, the highest frequency of the nerve block profile was found in cutaneous HFS and in the menthol model.

QST profiles compatible with primary hyperalgesia increased from 380 (58%) in the deterministic to 470 (72%) in the probabilistic algorithm. Beyond the defining models, this profile was frequent in the area of secondary hyperalgesia of the UVB radiation model, topical menthol and muscular HFS.

QST profiles consistent with secondary hyperalgesia were more than twice as frequent in the probabilistic sorting ($n = 134$; 20%) that in the deterministic sorting ($n = 58$; 9%). Beyond the defining capsaicin injection and cutaneous HFS, this profile

was only found in the area of secondary hyperalgesia of topical capsaicin in a relevant frequency (deterministic: 27%, probabilistic: 65%).

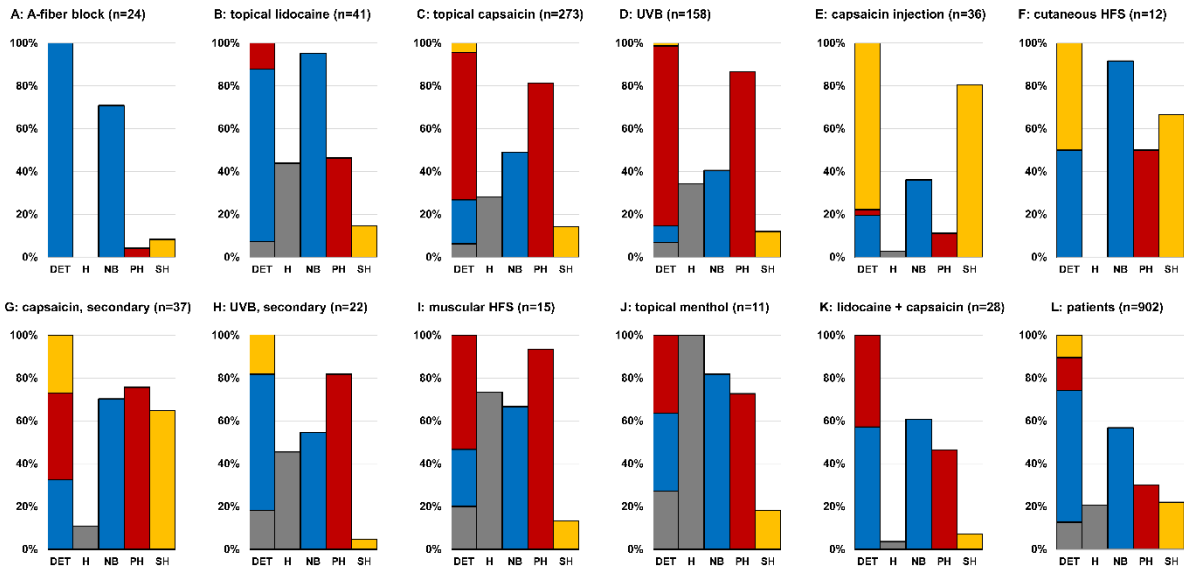


Figure 13: Deterministic and probabilistic sorting of human surrogate models and patients suffering from peripheral neuropathic pain. A: compression block (n=24), B: topical lidocaine (n=41), C: topical capsaicin (n=273), D: UVB sunburn (n=158), E: i.d. capsaicin injection (n=36), F: electrical high-frequency stimulation through punctate surface electrode (n=12), G: skin surrounding topical capsaicin (n=37), H: skin surrounding UVB sunburn (n=22), I: muscle HFS (n=15), J: topical menthol (n=11), K: topical lidocaine plus capsaicin (n=28), L: patients suffering from peripheral neuropathic pain (n=902). Deterministic sorting (DET): each profile is sorted to its best fit, probabilistic sorting: each profile is sorted to each mechanism for which it reaches a probability above 64%, as determined as optimal cut-off in a ROC analysis of patients suffering from neuropathic pain and healthy subjects in (Vollert et al., 2017b). H (grey): normal healthy skin, NB (Blue): nerve blocks, PH (red): primary hyperalgesia, SH (yellow): secondary hyperalgesia. From (Vollert et al., 2017b).

3.8 Patients in heuristic and mechanistic phenotypes

In 65% of the cases, patients sorted to the sensory loss phenotype were sorted to nerve block, patients sorted to the thermal hyperalgesia phenotype to the primary hyperalgesia pattern, and patients sorted to the mechanical hyperalgesia phenotype to the secondary hyperalgesia pattern (Table 14). This corresponded to a Cohen's kappa coefficient of agreement (scale: 0 = random classification, 1 = perfect

agreement between methods) of 0.44 (95% CI: 0.28 - 0.60), showing a substantial degree of agreement (Fleiss et al., 2003). The most prominent shifts were from the two hyperalgesia phenotypes in the heuristic patient clustering into the mechanistic phenotype consistent with nerve block.

In the probabilistic sorting (Table 14), interestingly, fewer patients were sorted to the nerve block profile (512 vs. 600 in deterministic sorting), while relatively more patients were sorted to primary (271 vs. 186 in deterministic sorting) or secondary hyperalgesia (198 vs. 116 in deterministic sorting). But the most prominent shift was still towards the profiles consistent with nerve blocks. These findings suggest the presence of multiple mechanisms in some of the patients. In fact, only one third of patients could be uniquely assigned to one of the three mechanisms. Another third of the patient QST profiles was consistent with multiple mechanisms ($n = 282$), most frequently a combination including the nerve block phenotype. The last third was not sufficiently distinct from a normal skin QST phenotype to be assigned to any mechanism ($n = 279$).

	sensory loss (n = 356)	thermal hyperalgesia (n = 267)	mechanical hyperalgesia (n = 279)
deterministic			
nerve block (NB, n = 600)	337	97	166
primary hyperalgesia (PH, n = 186)	6	156	24
secondary hyperalgesia (SH, n = 116)	13	14	89
probabilistic			
NB, n = 512	202	130	180
PH, n = 271	12	163	96
SH, n = 198	38	26	134
of the above: patients sorted to a single mechanism (n = 341)			
NB, n = 247	163	30	54
PH, n = 62	1	58	3
SH, n = 32	3	3	26
of the above: patients sorted to multiple mechanisms (n = 282)			
NB + PH (n=116)	4	84	28
NB + SH (n=73)	28	2	43
PH + SH (n=17)	0	7	10
NB + PH + SH (n = 76)	7	14	55
patients sorted to no mechanism (n= 279)			
	150	69	60

Table 14: Crosstabulation of dominant phenotype identified using the algorithm developed in patients (columns) vs. the algorithm developed based on surrogate models (rows) for n=902 patients. In the clear majority of the cases (65%, note that random assignment would be 33%) both algorithms concurred in the deterministic variant. From (Vollert et al., 2017b).

Figure 14 illustrates the similarity in QST profiles grouped either according to the original cluster analysis or the mechanistic sorting. While the profiles of patients sorted according to both algorithms are highly similar, the profiles of patients and surrogate models in the mechanistic sorting displays some interesting differences (compare Fig. 12 with Fig. 14B): patients sorted to the nerve block suffer from hypoalgesia, while in surrogate models pain thresholds are normal or show mild gain. Patients sorted to nerve block also have loss of VDT, which could not be modeled in the A-fiber block. For primary hyperalgesia, the models display almost no cold, but an isolated heat hyperalgesia, while patients show both cold and heat hyperalgesia. For secondary hyperalgesia, participants under surrogate models have more gain in MPT than in MPS, which is reversed in patients.

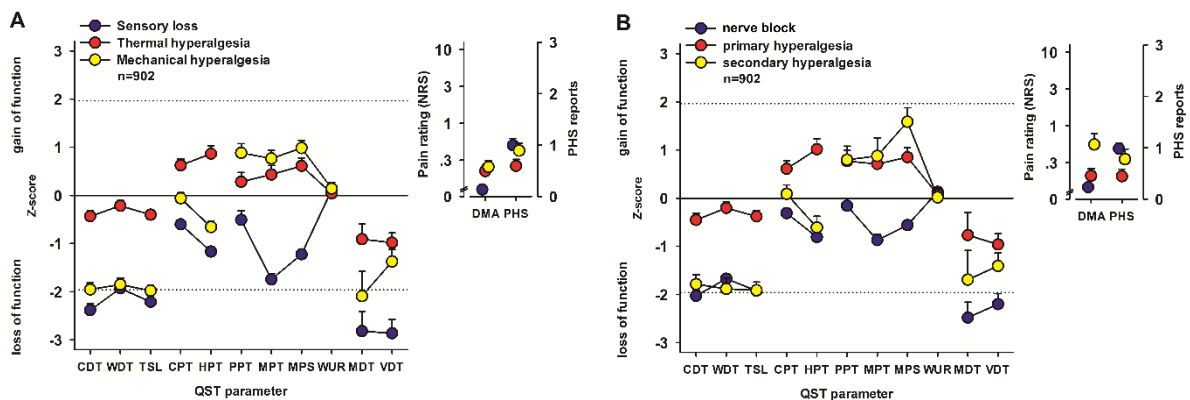


Figure 14: Average profiles per mechanism-stratified patient group. Using the deterministic algorithm, 902 patients with peripheral neuropathic pain (Baron et al., 2017) were sorted either according to three heuristic patient clusters (A) or according to three mechanistic clusters from human surrogate models (B). Positive z-scores indicate positive sensory signs (hyperalgesia), negative z-values indicate negative sensory signs (hypoesthesia, hypoalgesia). Dashed lines: 95% confidence interval for healthy participants ($-1.96 < z < +1.96$). Inserts show numeric pain ratings for DMA on a logarithmic scale (0-100) and frequency of PHS (0-3). From (Vollert et al., 2017b).

4 DISCUSSION

Aim of the presented work was to improve usefulness of Quantitative Sensory Testing for application in individual diagnostics, and to establish a mechanism-related classification of peripheral neuropathic pain that can be used for stratification of clinical trials and for individual patients' therapy. This global aim, on which pain researchers work since over 25 years (Fields et al., 1998; Woolf et al., 1998; Baumgartner et al., 2002; Baron et al., 2012; von Hehn et al., 2012; Edwards et al., 2016), has been progressed to a substantial degree. The main findings of this work were:

As a premise to using QST in multicenter clinical trials, it is crucial that examiners across centers, countries and languages produce comparable results. In an analysis of heterogeneity of patient and healthy participant data from ten European centers, eight countries and eight languages, it could be shown that skilled QST examiners are able to produce homogeneous results (Vollert et al., 2016a). The I^2 index of heterogeneity, ranging from 0% (no heterogeneity) to 100% (completely heterogeneous data) (Higgins and Thompson, 2002) was found to be 0% for all QST parameters for patients suffering from painful polyneuropathy or peripheral nerve injury, and 0% for healthy participants for all QST parameters except MDT (negligible 5.4%) and PPT (substantial 41.8%). Notably, the heterogeneity is lower in patients, where it is crucial that centers test reproducibly.

In a cluster analysis in 902 patients suffering from peripheral neuropathic pain, it could be shown that three sensory phenotypes appear across etiologies, and are stable across cohorts, as they are validated in a second cohort of 233 patients. These phenotypes are mainly characterized by

1. Hypoesthesia, both thermal and mechanical. This phenotype is labelled "sensory loss",
2. Comparably intact sensory function, often combined with thermal hyperalgesia or allodynia. This phenotype is labelled "thermal hyperalgesia" and
3. Thermal hypoesthesia, combined with mechanical hyperalgesia or allodynia, therefore called "mechanical hyperalgesia".

An algorithm that is able to assign single patients to each of these phenotypes was developed, validated in the original cohort and tested for its capability to separate

healthy participants from patients with neuropathic pain. Using this algorithm, sample sizes of patient groups that need to be screened to identify a phenotype-stratified subgroup large enough to conduct a clinical trial were calculated. These numbers suggest that crossover design studies are realistic for all phenotypes and the etiologies diabetic polyneuropathy, peripheral nerve injury and post-herpetic neuralgia, while parallel design studies will need very large screened populations for many phenotype-etiology combinations.

To further validate mechanistic basics of the identified phenotypes, a cluster analysis with identical protocol was conducted in $n = 657$ QSTs of healthy participants under human surrogate models of neuropathic pain. While the cluster analysis revealed no sufficient results, a sorting algorithm previously validated for sensory profiles of neuropathic pain patients (Vollert et al., 2017a) led to reproducible sorting of surrogate model sensory profiles into patterns defined a priori according to known mechanisms in a randomized split half analysis. Sorting of 902 neuropathic pain patients into these mechanistic phenotypes led to a similar distribution as the original heuristic clustering of the same patients (Baron et al., 2017; Vollert et al., 2017b; Vollert et al., 2017a).

4.1 Heterogeneity between centers

[The following section has been taken in parts and modified from (Vollert et al., 2016a).]

A central goal of the consortia DFNS, IMI Europain and Neuropain was to jointly gather data of patients suffering from neuropathy from centers across Europe, assessed in a highly valid protocol for a psychophysical testing procedure with broad scientific and clinical applications (Rolke et al., 2006a; Backonja et al., 2013) in a central database. One essential premise for performing multi-center analyses is homogeneous, unbiased results across all centers. The presented analysis of heterogeneity showed that there is no evidence of systematic heterogeneity between the QST assessment of the ten included centers located in eight countries across Europe and that the data can be analyzed as a homogenous group of patients.

The fact that these centers were able to produce highly reliable results in QST assessment, provided that the examiners are well trained and a strict protocol is applied, corresponds to previous findings in healthy participants (Rolke et al., 2006a;

Geber et al., 2009; Krumova et al., 2012a; Vollert et al., 2015) and patients suffering from painful neuropathies (Felix and Widerstrom-Noga, 2009; Geber et al., 2011). Still, this was the first multi-national and multi-lingual analysis of heterogeneity, showing that reproducible results can be achieved in bigger international consortia.

These findings are in line with a previous study reporting high test-retest and inter-observer reliability of QST in accordance with the DFNS protocol and training (Geber et al., 2011). QST has been suspected to be unreliable due to its subjective rating character (Yarnitsky and Granot, 2006), but it has been shown that subjective pain ratings correlate with neural activity in cortical regions associated with pain processing (Coghill and Eisenach, 2003). Furthermore, seemingly objective measures like corneal confocal microscopy or gold standards like skin biopsy for determining the intraepidermal nerve fiber density for detecting small fiber neuropathy have been conflictingly reported to be of high (Smith et al., 2005; Pacaud et al., 2015) or rather moderate (Wopking et al., 2009; Hertz et al., 2011) reliability, indicating that standardized procedures and extensive quality control are of higher importance for reliability than the method itself.

These results hold direct implications on the future planning for multi-center studies and clinical trials. To our knowledge, these have been avoided since the homogeneity of QST results across different labs has been previously doubted. However, it seems to be possible to produce highly reliable DFNS QST results across different research units, and, more importantly, also across countries and languages. The DFNS protocol is applied by trained examiners meanwhile in more than 70 research units in 21 countries across the world, and it is a future challenge to show that QST data from other continents beyond Europe is comparable to the reference data. A first indication is a study by Haroun et al. (unpublished to this point), in which a well-trained examiner performed DFNS QST in 52 healthy participants in Mumbai, India, without systematic deviations from the DFNS reference data.

There are no center effects that can be assigned to systematic heterogeneity in a meaningful way, and the I^2 Index for heterogeneity between centers was found to be above 0% only for the mechanical detection threshold (5.4%) and pressure pain threshold (41.8%) in healthy participants. While 5% for MDT can safely be assigned to normal scatter in (comparably) small samples, 42% for PPT can be considered as

high. This value was mainly influenced by two centers from Denmark and Sweden, but due to the broad confidence intervals of the center-specific means, the confidence interval of I^2 is still 0% to 66%, and Higgins and Thompson state that an I^2 Index with a confidence interval that includes 0% can be considered as no systematic heterogeneity (Higgins and Thompson, 2002). An indication of true heterogeneity in the assessment between the centers would be if the deviation from the overall mean in PPT of the above mentioned two centers was found also in the polyneuropathy and peripheral nerve injury group. This is only the case for one center in the polyneuropathy group, while the other three center-specific means (one center in the PNP group and both centers in the PNI group) lie very closely to the overall mean of PPT in the patient groups. Nonetheless, the high heterogeneity in PPT may be explained by the methodological peculiarities of the PPT assessment. This parameter, which assesses the only deep pain threshold in the protocol, is harder to be performed in the standardized way as per protocol in comparison to the other QST parameters, especially regarding the speed of the stimulus application, as it is the only stimulus ramp in the DFNS protocol that is controlled manually (Rolke et al., 2005; Mainka et al., 2014). Therefore, it is plausible that we found heterogeneity only in this parameter. The effect might be reduced if a device with a defined stimulus ramp application would be used (Mainka et al., 2014), still, this would further increase costs of the DFNS protocol. The effects are smaller over muscle than over bone structure (Mainka et al., 2014), therefore, the PPT is tested over muscle.

One of the main drawbacks of the analysis is the small group sizes per center. While in healthy participants a normal range is well defined for DFNS QST (Rolke et al., 2006a; Magerl et al., 2010; Pfau et al., 2014), however, patients suffering from painful neuropathies present with broadly scattering QST profiles (Maier et al., 2010; Backonja et al., 2013). It is important to keep in mind that the fact that CIs and I^2 indices are not higher for patients in comparison to healthy participants does not imply that heterogeneity between patients is comparable to heterogeneity between healthy participants. Both I^2 and the forest plot indicate measures of heterogeneity between centers, not between subjects. Individual differences can be influenced by many factors, such as underlying disease and its duration (Maier et al., 2010; Krumova et al., 2012a; Backonja et al., 2013). A high degree of individual scatter can lead to broad confidence intervals of the center-specific means, which may lead to underestimating the true heterogeneity (false negative results). On the other hand,

this may lead to a high degree of heterogeneity that is in fact not due to differing application of the QST protocol but different patient groups in phenotype of pathology or disease duration (false positive results).

While we can rule out the latter in our case, as we found no heterogeneity in the patient groups, the former is more difficult to conclude on. In 50% (11 of 22) of the QST parameters in the patient groups, the upper boundary of the 95% confidence interval of the I^2 Index is above 60%, indicating that although all I^2 values are found to be 0%, we find a very different picture if we look at the upper boundaries of the confidence intervals. Center-specific mean z-values scatter broadly for many QST parameters in the patient groups (e.g. mechanical pain threshold), and the I^2 values are low only because of the broad confidence intervals of the center-specific means. This emphasizes that our results should be treated with some caution, and analyses in larger patient groups may produce more nuanced results. Broad individual scatter within and between the centers in combination with small sample sizes can also result in the model finding no center-specific mean (Self and Liang, 1987; Andrews, 1997), as it has been the case for wind-up ratio and vibration detection threshold in patients suffering from peripheral nerve injury. This should not be interpreted as a definitive lack of heterogeneity between the centers, but merely reflect that on the basis of the underlying data we cannot make any clear statements about heterogeneity for these parameters in the PNI group.

4.2 Sensory phenotypes in patients

[The following section has been taken in parts and modified from (Baron et al., 2017) and (Vollert et al., 2017a).]

In a cluster analysis of $n = 902$ patients suffering from peripheral neuropathic pain without a predefined number of clusters, we found that a three-cluster-solution best describes patients with peripheral neuropathic pain. All subgroups occurred in relevant numbers across etiologies, but frequencies differed between the entities. This 3-cluster solution and the structure of the sensory profiles could be reproduced in the validation cohort. It matches the three subgroups described in smaller studies in patients with PHN almost 20 years ago (Fields et al., 1998). The subsequently developed algorithm enables individual allocation of patients to one or more sensory

phenotypes, and, further, can separate patients from healthy participants with a very high specificity of 94% and a sensitivity of 78%.

Although the presented algorithm offers a barrier for excluding healthy participants, this should be considered within the clinical context. We decided on a rather conservative criterion with high specificity. It should be noted that patients eligible for clinical trials are usually screened beforehand, have been shown to have a lesion or disease and have spontaneous pain. Confirming neuropathic pain is relying on a grading system, following four steps:

1. a history of a relevant neurological lesion or disease,
2. anatomically plausible pain distribution
3. sensory signs of neuropathy, and finally
4. diagnostic tests confirming the lesion or disease (Finnerup et al., 2016).

Our algorithm serves as step (3) for confirming or dismissing neuropathic pain, but on an averaged level across a full profile: e.g., a strongly decreased vibration or thermal detection in an otherwise normal profile would be considered a single negative sensory sign, but might still result in a high averaged probability of being healthy in this algorithm. Further studies will provide insights, if single negative sensory signs or a phenotype probability is more meaningful. A second use of our algorithm is for stratification of patients suffering from neuropathic pain according to pain phenotype, and its efficacy will have to be validated in future RCTs.

4.2.1 Sensory loss phenotype

Cluster 1 (42%) was characterized by a loss of small and large fiber function and the presence of paradoxical heat sensations. These patients did not suffer from sensory gain except a mild dynamic mechanical allodynia in few patients. Roughly 52% of patients with polyneuropathy fell into this category indicating dying-back degeneration of nearly all fiber classes. Interestingly, 43% of patients with painful radiculopathy demonstrated this sensory pattern, suggesting severe degeneration of sensory fibers within the affected nerve root. Paradoxical heat sensation was most frequent, underpinning the notion that it is induced by a loss of afferent input although at face value it is a positive sensory sign possibly related to a central disinhibition process (Yarnitsky and Ochoa, 1990; Hansen et al., 1996).

The sensory profile is similar to that of a compression nerve block (Fruhstorfer, 1984; Yarnitsky and Ochoa, 1991; Baumgartner et al., 2002). It likely represents the "deafferentation" or "painful hypoesthesia" subgroups described by others (Fields et al., 1998; Baumgartner et al., 2002; Truini et al., 2009; Hatem et al., 2010). The spontaneous pain was likely due to ectopic action potentials generated in proximal sites of injured nociceptors (Campbell and Meyer, 2006) e.g., in the dorsal root ganglion or in deafferented central nociceptive neurons (Devor et al., 1992; Orstavik et al., 2006; Serra et al., 2012).

4.2.2 Thermal hyperalgesia phenotype

Cluster 2 was characterized by relatively preserved large and small fiber sensory function in combination with heat and cold hyperalgesia and only low intensity dynamic mechanical allodynia. Burning pain quality in this cluster was less prominent than in the other groups, consistent with findings in Guillain-Barré syndrome where burning pain was associated with small fiber deficits (Martinez et al., 2010). This pattern occurred in 33% of all peripheral neuropathic pain patients regardless of etiology. The fact that in one third of all patients the cutaneous sensory function was relatively preserved despite documented nerve damage indicates that peripheral neuropathic pain may be associated with effective cutaneous regeneration and sensitized nociceptors.

The sensory profile is similar to that of a UVB burn lesion (Gustorff et al., 2013) and is likely due to peripheral sensitization (Treede et al., 1992). It represents the "irritable nociceptor" subgroup described by others (Fields et al., 1998; Ochoa et al., 2005; Demant et al., 2014; Demant et al., 2015). Sensitized nociceptors are associated with overexpression of channels and receptors leading to pathological spontaneous discharges and a lowered activation threshold for thermal (heat and cold) and mechanical stimuli. Ongoing hyperactivity in surviving nociceptors may be responsible for ongoing pain (Campbell and Meyer, 2006) and may lead to some central sensitization in the spinal cord dorsal horn, so that tactile stimuli conveyed in A-fibers become capable of activating central nociceptive neurons. As a result, mechanical stimuli induce enhanced pain percepts, i.e., pinprick hyperalgesia and dynamic mechanical allodynia (von Hehn et al., 2012). Since these types of mechanical hyperalgesia were only present in about 20% of the patients, peripheral

nociceptor drive obviously does not always induce central sensitization (Truini et al., 2013).

4.2.3 Mechanical hyperalgesia phenotype

Cluster 3 (24%) was characterized by a predominant loss of cold- and heat-sensitive small fiber function in combination with blunt pressure hyperalgesia, pinprick hyperalgesia and more frequent dynamic mechanical allodynia. The profile was most commonly present in patients with PHN (47%). It is similar to the one induced by high-frequency electrical stimulation of the skin that is capable of inducing spinal long-term potentiation (Randic et al., 1993; Lang et al., 2007) and likely equivalent to "neurogenic hyperalgesia" or "central sensitization" subgroups described by others (Fields et al., 1998; Baumgartner et al., 2002). Central sensitization is prominent for mechanical stimuli (Baumann et al., 1991; Simone et al., 1991; Treede et al., 1992) but not thermal stimuli. The dissociation of thermal and mechanical hyperalgesia may be explained by differences in neural signaling of thermal and mechanical pain that starts with peripheral encoding in distinct subsets of nociceptors (Cavanaugh et al., 2009; Henrich et al., 2015). Ongoing pain in this subgroup indicates spontaneous activity in the nociceptive system, which may originate in the peripheral and/or central nervous system.

4.2.4 Sample size recommendations

While a series of studies showed that a post-hoc responder analysis can reveal phenotypes that are important to predict treatment response (Attal et al., 2004; Wasner et al., 2005; Edwards et al., 2006; Simpson et al., 2010; Westermann et al., 2012; Katz et al., 2015; Attal et al., 2016; Mainka et al., 2016; Reimer et al., 2017), the first phenotype-stratified, randomized, placebo-controlled trials have been published only recently (Demant et al., 2014; Demant et al., 2015). In these studies, oxcarbazepine showed a superior effect over placebo in a subgroup with "irritable nociceptors", a group with a sensory profile very similar to the thermal hyperalgesia phenotype in this study. In contrast, for topical lidocaine no group difference could be demonstrated.

The main problem with the definition of "irritable nociceptors" based on individually abnormal QST values and loss and gain of functions patterns (Rolke et al., 2006a;

Maier et al., 2010) is that it is based on a statistically very sound, but strictly conservative approach with (comparably) low sensitivity. For instance in diabetic polyneuropathy the “irritable nociceptor” phenotype is virtually not present (Themistocleous et al., 2016).

The approach taken in this study does, in contrast, not rely on individually abnormal QST values, but focusses on cluster centroids. An appealing advantage of this dynamic method can be seen in Figure 9. The thermal hyperalgesia phenotype, which is similar to the “irritable nociceptor” and may have similar underlying mechanisms of pain generation, is found to be a prominent phenotype in a reasonable subgroup of roughly one third of the patients with neuropathic pain due to diabetic polyneuropathy. In a cohort of patients suffering from PHN, PNP or PNI treated with oxcarbazepine (Demant et al., 2014), the thermal hyperalgesia phenotype is similarly effective as predictor of treatment efficacy compared to phenotyping by the “irritable nociceptor” classification.

To use stratification into subgroups in clinical trials a large patient population must be screened beforehand. As screening with QST may be considered expensive and time-consuming, not only needs the final stratified study population be considered, but also the number of patients necessary to screen. Thus, a solid sample size calculation is a prerequisite for a stratified study. In Table 9 we present sample size numbers for screening of populations for painful diabetic polyneuropathy, painful peripheral nerve injury and post-herpetic neuralgia in relation to estimated effect size (0.3 vs. 0.5 vs. 0.7) and study design (parallel vs. crossover). Crossover sample sizes are overall “realistic” numbers – across phenotypes and clinical etiology based entities. If a parallel study design is intended, however, phenotype stratification may only be possible if a high effect size (e.g. 0.7) is anticipated. Another way to deal with high numbers to screen would be to limit the screening process to few centers, while the study main RCT can be conducted after screening at multiple sites.

Of the two methods of sorting patients to phenotypes presented in parallel here, we do not recommend one or the other in general, because they have individual advantages and disadvantages. The deterministic approach, sorting each patient to exactly one phenotype, ignores that multiple mechanisms of pathology may be present in a patient, and that these mechanisms may overlay each other and result in a sensory phenotype that cannot easily be allocated to one phenotype over the other.

The probabilistic approach, however, holds its own challenges: while the overlap between phenotypes is reasonable for peripheral nerve injury and post-herpetic neuralgia, patients with diabetic polyneuropathy tend to present more than one phenotype with significant probability. This effect, probably caused by the overwhelming effect of loss symptoms in these patients, may dilute especially the mechanical hyperalgesia phenotype, and to a lesser extent the thermal hyperalgesia phenotype. When screening for these phenotypes in patients with diabetic polyneuropathy, this limitation should be considered by rather using the deterministic algorithm (or a probabilistic assignment on mechanistic basis developed in participants under surrogate models). For peripheral nerve injury and post-herpetic neuralgia, a notable part of the patients (21% and 30%, respectively) is not sorted to any phenotype in the probabilistic algorithm and therefore excluded from the analysis. While this is acceptable for phenotype-stratified trials, it becomes a problem if the algorithm would be used for designing individual patients' treatment strategy in the future. Again, the deterministic approach might be favorable in this case.

4.3 Subgrouping human surrogate models

[The following section has been taken in parts and modified from (Vollert et al., 2017b).]

While the clustering approach for human surrogate models without a-priori assumptions did not lead to a sufficient result, a mechanistically guided grouping of models into surrogates for denervation, peripheral sensitization and central sensitization led to three mechanistic phenotypes that show high similarity to the phenotypes heuristically found in patients.

4.3.1 Nerve blocks as human surrogate model of denervation

The sensory profile of human surrogate models for denervation (i.e. compression block and topical lidocaine) was characterized by pronounced loss in thermal and mechanical detection thresholds in combination with paradoxical heat sensations. These patterns have been reported in previous single center studies (Ziegler, 1999; Klein et al., 2005; Krumova et al., 2012b) and may hence serve to validate our multi center QST profiling approach. Consistent with its selective effect on myelinated nerve fibers, compression nerve block had larger effects on mechanical and cold

detection thresholds than on warm detection threshold (Ziegler, 1999; Henrich et al., 2015). Paradoxical perception of cooling stimuli as warm has been reported by some but not all of these studies (Fruhstorfer, 1984). Pinprick stimuli are often perceived as less painful in complete A-fiber block (Ziegler, 1999; Henrich et al., 2015), while mild pinprick hyperalgesia has been found associated with preserved A-delta fiber function (Andrew and Greenspan, 1999; Jørum et al., 2000). Topical lidocaine had mild effects when compared to complete conduction block by regional anesthesia (Gandevia and Phegan, 1999; Klein et al., 2005). Other clinical trials using this lidocaine patch did not assess sensory profiling.

4.3.2 Primary hyperalgesia as human surrogate model of peripheral sensitization

The profile of surrogate models of peripheral sensitization (topical capsaicin or UVB radiation) was characterized by pronounced hyperalgesia to heat, pressure and pinprick pain, but also presented very mild thermal sensory loss. Primary nociceptive afferents are easily sensitized to heat stimuli, but much less so to von Frey or pinprick stimuli (Treede et al., 1992). Peripheral sensitization to heat may be explained by phosphorylation of the heat-gated cation channel TRPV1 through multiple pathways (Voets et al., 2004), but has also been shown to be induced by TRPA1 agonist allyl-isothiocyanate (Andersen et al., 2017). While some peripheral sensitization to blunt pressure has been reported before (Kilo et al., 1994), pinprick hyperalgesia in these models may indicate additional central sensitization induced by enhanced peripheral nociceptive input to the spinal cord. Sensory loss occurred mostly in the topical capsaicin model and was restricted to thermal detection thresholds; this likely reflects desensitization by the TRPV1 agonist capsaicin, which is the intended clinical mode of action (Hayman and Kam, 2008).

4.3.3 Secondary hyperalgesia as human surrogate model of central sensitization

Human surrogate models of central sensitization (intra-dermal capsaicin and electrical high-frequency stimulation) were characterized by pronounced pinprick hyperalgesia, but also pronounced thermal sensory loss. Other human surrogate models using intra-cutaneous electrical stimulation have not yet undergone full sensory profiling, but published data are also consistent with a secondary hyperalgesia model (Koppert et al., 2001). Combined studies in monkey and humans using i.d. capsaicin have

shown pronounced increases in both WDT and high threshold spinal neuron output despite unchanged A- and C-nociceptor input (Baumann et al., 1991; Simone et al., 1991). The same studies suggested that dynamic mechanical allodynia is a hallmark sign of central sensitization. Hyperalgesia to blunt pressure was mild in these models as compared to the primary hyperalgesia models, suggesting that this type of mechanical hyperalgesia may be primarily peripherally mediated (Ochoa, 1993; Koltzenburg et al., 1994). Tactile sensory loss has been reported in human surrogate models of secondary hyperalgesia (Magerl and Treede, 2004) and also in patients with pinprick hyperalgesia (Geber et al., 2013). An inverse spinal gate with small fiber input inhibiting processing of large fiber input (Zimmermann, 1968; Mendell, 2014) has been suggested as a mechanism. Thermal sensory loss is a new finding, suggesting that broad loss of detection of non-painful stimuli may be a characteristic feature of central sensitization (according to the IASP definition, this includes contributions by descending control systems). This implies that sensory loss in chronic pain patients does not necessarily have to be due to structural changes (e.g. intraepidermal nerve fiber loss), but may also be a functional sign, and hence potentially sensitive to analgesic treatment regimes.

Additionally, the secondary hyperalgesia phenotype would have been suspected in the skin surrounding a primary hyperalgesia area induced by topical capsaicin or UVB irradiation. Sensory profiles (Fig. 11 and Table 10) confirmed this prediction, albeit in mild form. This may explain why the sorting algorithm assigned only 63% (skin surrounding topical capsaicin) and 4% (surrounding UVB) to the secondary hyperalgesia phenotype.

4.3.4 Other human surrogate models

The sensory profile of topical lidocaine plus topical capsaicin displays a combination profile of nerve block and primary hyperalgesia, but is unlike the secondary hyperalgesia phenotype. In the cluster analysis of neuropathic pain patients (Baron et al., 2017), the heuristic “mechanical hyperalgesia” phenotype exhibited a combination of sensory loss and hyperalgesia; one might have expected that this mixture could be induced by combining an experimental nerve block with an experimental hyperalgesia model. The present analysis indicates that primary and secondary hyperalgesia models are still distinguishable in the presence of a

concomitant mild nerve block. This finding suggests that sensory profiling may also be able to distinguish between contributions of peripheral vs. central sensitization in patients with neuropathic pain that have concomitant sensory loss.

4.3.5 Mechanistic significance of heuristically found phenotypes

QST data for both patients and human surrogate models were normalized to the same published reference data. Therefore, we could immediately apply the sorting according to surrogate model profiles to our previously published patient data. Sorting of 902 neuropathic pain patients into mechanistic phenotypes led to a similar distribution as the original heuristic clustering (65%, Cohen's $\kappa=0.44$, note that the expected value of random sorting is 33%). This supports previous mechanistic interpretations of the clinically found phenotypes: the thermal hyperalgesia patient phenotype shows strong overlap with surrogate models of primary hyperalgesia. Both show no pronounced loss of thermal detection, which indicates intact small fiber function in both patients and models. This supports previous interpretations as irritable nociceptor (Fields et al., 1998) and peripheral sensitization (Truini et al., 2013). Both evoked and ongoing pain is likely to be due to surviving nociceptors in these patients.

The mechanical hyperalgesia patient phenotype shows a strong overlap with surrogate models of secondary hyperalgesia, which supports an interpretation of this phenotype to be a phenotype of reorganization or central sensitization. Substantial thermal sensory loss (indicating loss of small fibers or small fiber function) suggests that also damaged nociceptors are involved, generating ongoing pain and inducing central sensitization (Baron et al., 2013).

The sensory loss patient phenotype shows a strong overlap with experimental nerve blocks. These blocks were frequently used as tools to identify normal sensory function of fiber classes (A vs. C), but not yet widely recognized as mimicking aspects of neuropathic pain (Baumgartner et al., 2002; Klein et al., 2005). Both the clinical phenotype and the surrogate models are dominated by loss of small and large fibers or fiber function. This supports an interpretation as denervation or deafferentation, where central neurons may develop denervation super-sensitivity to other inputs (Colloca et al., 2017). Mild pinprick hyperalgesia appears to be intrinsic to these mechanisms.

While this relationship of patient phenotype and mechanistically defined surrogate model phenotype was true for about 2/3 of the patients, most of the remaining third were assigned to the nerve block phenotype according to the surrogate models vs. one of the hyperalgesia phenotypes in the heuristic patient cluster analysis. This systematic shift is consistent with the presence of partial nerve damage in most of the neuropathic pain conditions that may lead to coexistence of denervation and sensitization of the remaining pathways (Campbell and Meyer, 2006). In fact, in the probabilistic sorting algorithm, roughly one third of the patients had evidence for multiple mechanisms, most frequently a combination with the denervation phenotype. Of interest, the primary hyperalgesia phenotype was more frequent than the secondary hyperalgesia phenotype, suggesting that the relevance of peripheral sensitization of surviving nociceptors may have been underestimated in the past. In turn, central sensitization may be more frequent than the secondary hyperalgesia phenotype, since it may also be induced by ectopically generated impulses from damaged nociceptors or by enhanced input from sensitized nociceptors. Based on these findings, the probabilistic sorting may be a useful approach for mechanism-based patient stratification. These data cover, however, only peripheral neuropathic pain, and to this point we cannot make any extrapolations onto, e.g., central pain, nociceptive pain, or deep pains.

4.4 Limitations

[The following section has been taken in parts and modified from (Vollert et al., 2016a), (Baron et al., 2017), (Vollert et al., 2017a) and (Vollert et al., 2017b).]

This analysis is based on QST z-values, which are dependent on the availability and quality of the underlying normative data, which are constantly in development (Rolke et al., 2006a; Blankenburg et al., 2010; Magerl et al., 2010; Pfau et al., 2014). There are methodological limits in the assessment, e.g., for the cold pain threshold at the feet, there is only a small window for abnormal values narrowing with growing age (Rolke et al., 2006a) (compare abnormal findings for CPT in patients with polyneuropathy), some sharp mechanical stimuli that are part of the standard protocol are too painful if applied in sensitive body areas as the cheek (Rolke et al., 2006a) or in children (Blankenburg et al., 2010). This is especially a problem in assessing the wind-up ratio, as for an important part of the patients, either the

sharpest stimulus (512nm) is rated as not painful (and a ratio of zero cannot be calculated), or the softest stimulus is rated as too painful to endure a series of ten stimuli. This leads to a missing rate of WUR of over 10% of the patients (Baron et al., 2017).

Since the inclusion criteria slightly differed between the three consortia, there is no perfect homogeneity of patients' clinical history within etiologies. Furthermore, in contrast to short-term stability of QST long-term stability over weeks has not been studied largely and is known to change over time during major events like adolescence (Hirschfeld et al., 2012). Hence, it is unknown whether patients can shift from one cluster into another. Patients may, however, suffer from more than one type of mechanism which changes during the course of chronicity.

The sample size calculations in Table 9 show both advantage and disadvantage of a QST-based phenotype stratification for clinical trials. A novel drug that is aiming at a phenotype that is only present in a fifth of the population will never show an effect superior to placebo in a non-stratified population. On the other hand, many patients must be screened to identify an eligible subpopulation, and screening with QST needs substantial training to be reliable. Furthermore, some QST parameters are mechanistically linked and therefore probably intercorrelated (e.g., CDT or WDT and TSL). In the presented algorithm, these domains will be slightly overweighed. While beyond the scope of this thesis, a factorial analysis of the QST protocol is one of the upcoming tasks to show the importance and meaning of each parameter in relation to the full protocol.

The human surrogate models entered into our analysis do not explicitly cover the actions of endogenous pain modulating systems (Kennedy et al., 2016), which may also play an important clinical role in pain generation. Descending modulation may contribute to the secondary hyperalgesia phenotype, but might also exhibit yet another sensory profile. This implies that the list of distinguishable sensory phenotypes may be longer than the three we describe, but not shorter.

In many cases, surrogate models were tested in proximal upper or lower limbs, for which no separate published reference data are available (Rolke et al., 2006a; Magerl et al., 2010; Pfau et al., 2014). This may have confounded our z-values. To control for this factor, we compared effect sizes of treatment vs. untreated sites (same area before treatment or contralateral, untreated side) to effects on z-values.

These data showed that effect sizes and z-values generally concur, and that the confounding element of non-reference areas is neglectable. One exception should be noted: secondary hyperalgesia models (i.d. capsaicin and cutaneous HFS) show significant loss of thermal detection on the z-scale, while effect sizes compared to the untreated area are present, but smaller (0.2 – 0.7). Loss of thermal detection might therefore be overestimated in the secondary hyperalgesia mechanism. Still, it should be noted that patients are often tested beyond reference areas as well, so this limitation partly reflects a normal clinical setting.

An additional limitation for the use in large clinical trials is that the QST protocol requires substantial training and an expensive device. In the long run, both for large trials and daily clinical practice, an approximation via a simple bed-side testing protocol would be highly valuable.

4.5 Impact and conclusions

[The following section has been taken in parts and modified from (Vollert et al., 2016a), (Baron et al., 2017), (Vollert et al., 2017a) and (Vollert et al., 2017b).]

In summary, this work showed that

1. centers across Europe can produce similar results when assessing patients' and participants' sensory profile,
2. profiles of patients suffering from peripheral neuropathic pain persistently show three sensory subgroups, which are present across etiologies, genders and age decades,
3. a simple algorithm can be used for stratification of patients into these three sensory phenotypes for use in clinical trials and in the future to indicate individual patients' optimal treatment strategies,
4. frequencies of these phenotypes differ between etiologies of neuropathic pain, which should affect the number of patients screened for clinical trials,
5. similar phenotypes are identified based on mechanistic assumptions in human surrogate models of neuropathic pain.

The parts of this work that are already published were well accepted in the pain research community (Dworkin and Edwards, 2017). As a result of this analysis, the European Medicines Agency's (EMA) committee for medicinal products for human

use recommends the phenotype stratification presented here for determining eligible sensory phenotypes of patients in exploratory trials on neuropathic pain, as also incorporated in the new EMA guideline for clinical development of new treatments for pain (European Medicines Agency, 2016). Using the probabilistic sorting according to human surrogate model profiles, patients suffering from neuropathic pain can be tentatively stratified in future studies to presumed underlying mechanisms. It should be noted, however, that the three classes of human surrogate models studied here likely represent combined rather than single mechanisms (e.g. peripheral and central sensitization in primary hyperalgesia). This, however, is likely true also for studies in awake behaving animals. Therefore, a reverse translation approach may be useful for developing novel analgesic medications, if they are initially validated in animal models of nerve block, primary or secondary hyperalgesia. Medications effective on these phenotypes can easily be validated in human surrogate models and then transferred to subgroups of neuropathic pain patients.

5 SUMMARY

Quantitative Sensory Testing (QST) following the DFNS (German Research Network on Neuropathic Pain) protocol assesses the function of the somatosensory nervous system. Long-term aim of QST research is the establishment of a mechanism-based classification of neuropathic pain. Over the last years, a central database with QST assessments of healthy participants, healthy participants under human surrogate models of neuropathic pain, and patients suffering from neuropathic pain has been built within the European consortia IMI Europain, Neuropain and the DFNS.

Aim of this work was to show that QST assessment is comparable between the participating centers across Europe in an analysis of heterogeneity, to use unsupervised clustering methods to identify subgroups of sensory profiles appearing across etiologies of peripheral neuropathic pain and may indicate underlying mechanisms of pathophysiology, to develop an individual assignment algorithm sorting QST profiles to these subgroups, to estimate the frequency of these subgroups across the common entities of peripheral neuropathic pain diabetic polyneuropathy, peripheral nerve injury and post-herpetic neuralgia and to further validate the subgroups identified in patients in surrogate models of neuropathic pain, in which the underlying mechanisms are well described.

Heterogeneity was overall low between the 11 participating European centers and the 13 QST parameters for healthy participants, and virtually non-existing for patients suffering from polyneuropathy or peripheral nerve injury. The cluster analysis found three sensory phenotypes, which are mainly characterized by either sensory loss (SL), intact sensory function and mild thermal hyperalgesia (TH) or loss of thermal detection and mild mechanical hyperalgesia (MH). The most common phenotype in diabetic polyneuropathy was SL (83%), followed by MH (75%) and TH (34%, note that percentages are overlapping and not additive). In peripheral nerve injury, frequencies were 37%, 59% and 50%, and in post-herpetic neuralgia, 31%, 63% and 46%. Surrogate models of nerve block were similar to the SL phenotype, but also showed mild pinprick hyperalgesia and paradoxical heat sensations. Peripheral sensitization models resembled the TH phenotype, while models of central sensitization showed high similarities to the MH phenotype.

These data suggest that classifying patients based on QST profiles in an approach developed hypothesis-free in patients and validated in models with well-described mechanisms may be a good strategy for mechanism-based stratification of neuropathic pain patients for future clinical trials, as encouraged by the European Medicines Agency EMA.

6 REFERENCES

Andersen, H.H., Lo Vecchio, S., Gazerani, P., and Arendt-Nielsen, L. (2017). A dose-response study of topical allyl-isothiocyanate (mustard oil) as human surrogate model of pain, hyperalgesia, and neurogenic inflammation. *Pain*.

Andrew, D., and Greenspan, J.D. (1999). Peripheral coding of tonic mechanical cutaneous pain: comparison of nociceptor activity in rat and human psychophysics. *Journal of neurophysiology* 82, 2641-2648.

Andrews, D. (1997). *Estimation When a Parameter Is on a Boundary: Theory and Applications*.

Attal, N., Andrade, D.C. de, Adam, F., Ranoux, D., Teixeira, M.J., Galhardoni, R., Raicher, I., Üçeyler, N., Sommer, C., and Bouhassira, D. (2016). Safety and efficacy of repeated injections of botulinum toxin A in peripheral neuropathic pain (BOTNEP). A randomised, double-blind, placebo-controlled trial. *The Lancet Neurology* 15, 555-565.

Attal, N., Rouaud, J., Brasseur, L., Chauvin, M., and Bouhassira, D. (2004). Systemic lidocaine in pain due to peripheral nerve injury and predictors of response. *Neurology* 62, 218-225.

Backonja, M.M., Attal, N., Baron, R., Bouhassira, D., Drangholt, M., Dyck, P.J., Edwards, R.R., Freeman, R., Gracely, R., and Haanpaa, M.H., et al. (2013). Value of quantitative sensory testing in neurological and pain disorders: NeuPSIG consensus. *Pain* 154, 1807-1819.

Baron, R., Förster, M., and Binder, A. (2012). Subgrouping of patients with neuropathic pain according to pain-related sensory abnormalities. A first step to a stratified treatment approach. *The Lancet Neurology* 11, 999-1005.

Baron, R., Hans, G., and Dickenson, A.H. (2013). Peripheral input and its importance for central sensitization. *Annals of neurology* 74, 630-636.

Baron, R., Maier, C., Attal, N., Binder, A., Bouhassira, D., Cruccu, G., Finnerup, N.B., Haanpaa, M., Hansson, P., and Hulleman, P., et al. (2017). Peripheral neuropathic pain: a mechanism-related organizing principle based on sensory profiles. *Pain* 158, 261-272.

- Baumann, T.K., Simone, D.A., Shain, C.N., and LaMotte, R.H. (1991). Neurogenic hyperalgesia: the search for the primary cutaneous afferent fibers that contribute to capsaicin-induced pain and hyperalgesia. *Journal of neurophysiology* 66, 212-227.
- Baumgartner, U., Magerl, W., Klein, T., Hopf, H.C., and Treede, R.D. (2002). Neurogenic hyperalgesia versus painful hypoalgesia: two distinct mechanisms of neuropathic pain. *Pain* 96, 141-151.
- Binder, A., Stengel, M., Klebe, O., Wasner, G., and Baron, R. (2011). Topical high-concentration (40%) menthol-somatosensory profile of a human surrogate pain model. *The journal of pain : official journal of the American Pain Society* 12, 764-773.
- Blankenburg, M., Boekens, H., Hechler, T., Maier, C., Krumova, E., Scherens, A., Magerl, W., Aksu, F., and Zernikow, B. (2010). Reference values for quantitative sensory testing in children and adolescents: developmental and gender differences of somatosensory perception. *Pain* 149, 76-88.
- Bouhassira, D., and Attal, N. (2016). Translational neuropathic pain research: A clinical perspective. *Neuroscience* 338, 27-35.
- Bouhassira, D., Attal, N., Fermanian, J., Alchaar, H., Gautron, M., Masquelier, E., Rostaing, S., Lanteri-Minet, M., Collin, E., and Grisart, J., et al. (2004). Development and validation of the Neuropathic Pain Symptom Inventory. *Pain* 108, 248-257.
- Campbell, J.N., and Meyer, R.A. (2006). Mechanisms of neuropathic pain. *Neuron* 52, 77-92.
- Cavanaugh, D.J., Lee, H., Lo, L., Shields, S.D., Zylka, M.J., Basbaum, A.I., and Anderson, D.J. (2009). Distinct subsets of unmyelinated primary sensory fibers mediate behavioral responses to noxious thermal and mechanical stimuli. *Proceedings of the National Academy of Sciences of the United States of America* 106, 9075-9080.
- Coghill, R.C., and Eisenach, J. (2003). Individual Differences in Pain Sensitivity. Implications for Treatment Decisions. *Anesthes* 98, 1312-1314.
- Cohen, J. (1992). A power primer. *Psychological Bulletin* 112, 155-159.
- Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences* (Taylor & Francis).

Colloca, L., Ludman, T., Bouhassira, D., Baron, R., Dickenson, A.H., Yarnitsky, D., Freeman, R., Truini, A., Attal, N., and Finnerup, N.B., et al. (2017). Neuropathic pain. *Nature Reviews Disease Primers* 3, 17002.

DeLong, E.R., DeLong, D.M., and Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44, 837-845.

Demant, D.T., Lund, K., Finnerup, N.B., Vollert, J., Maier, C., Segerdahl, M.S., Jensen, T.S., and Sindrup, S.H. (2015). Pain relief with lidocaine 5% patch in localized peripheral neuropathic pain in relation to pain phenotype: a randomised, double-blind, and placebo-controlled, phenotype panel study. *Pain* 156, 2234-2244.

Demant, D.T., Lund, K., Vollert, J., Maier, C., Segerdahl, M., Finnerup, N.B., Jensen, T.S., and Sindrup, S.H. (2014). The effect of oxcarbazepine in peripheral neuropathic pain depends on pain phenotype: a randomised, double-blind, placebo-controlled phenotype-stratified study. *Pain* 155, 2263-2273.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm.

Devor, M., Wall, P.D., and Catalan, N. (1992). Systemic lidocaine silences ectopic neuroma and DRG discharge without blocking nerve conduction. *Pain* 48, 261-268.

Dworkin, R.H., and Edwards, R.R. (2017). Phenotypes and treatment response. *Pain*, 1.

Dworkin, R.H., O'Connor, A.B., Backonja, M., Farrar, J.T., Finnerup, N.B., Jensen, T.S., Kalso, E.A., Loeser, J.D., Miaskowski, C., and Nurmikko, T.J., et al. (2007). Pharmacologic management of neuropathic pain: evidence-based recommendations. *Pain* 132, 237-251.

Edwards, R.R., Dworkin, R.H., Turk, D.C., Angst, M.S., Dionne, R., Freeman, R., Hansson, P., Haroutounian, S., Arendt-Nielsen, L., and Attal, N., et al. (2016). Patient phenotyping in clinical trials of chronic pain treatments: IMMPACT recommendations. *Pain* 157, 1851-1871.

Edwards, R.R., Haythornthwaite, J.A., Tella, P., Max, M.B., and Raja, S. (2006). Basal heat pain thresholds predict opioid analgesia in patients with postherpetic neuralgia. *Anesthesiology* 104, 1243-1248.

Enax-Krumova, E.K., Pohl, S., Westermann, A., and Maier, C. (2017). Ipsilateral and contralateral sensory changes in healthy subjects after experimentally induced concomitant sensitization and hypoesthesia. *BMC neurology* 17, 60.

European Medicines Agency (2016). EMA/CHMP/970057/2011. Guideline on the clinical development of medicinal products intended for the treatment of pain. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2016/12/WC500219131.pdf. 15.12.2016.

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3. A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 39, 175-191.

Felix, E.R., and Widerstrom-Noga, E.G. (2009). Reliability and validity of quantitative sensory testing in persons with spinal cord injury and neuropathic pain. *Journal of rehabilitation research and development* 46, 69-83.

Fields, H.L., Rowbotham, M., and Baron, R. (1998). Postherpetic neuralgia: irritable nociceptors and deafferentation. *Neurobiology of disease* 5, 209-227.

Finnerup, N.B., Attal, N., Haroutounian, S., McNicol, E., Baron, R., Dworkin, R.H., Gilron, I., Haanpaa, M., Hansson, P., and Jensen, T.S., et al. (2015). Pharmacotherapy for neuropathic pain in adults: a systematic review and meta-analysis. *The Lancet. Neurology* 14, 162-173.

Finnerup, N.B., Haroutounian, S., Kamerman, P., Baron, R., Bennett, D.L.H., Bouhassira, D., Cruccu, G., Freeman, R., Hansson, P., and Nurmikko, T., et al. (2016). Neuropathic pain: an updated grading system for research and clinical practice. *Pain* 157, 1599-1606.

Fleiss, J.L. (1973). *Statistical methods for rates and proportions* (New York: Wiley).

Fleiss, J.L., Levin, B., and Paik, M.C. (2003). *Statistical methods for rates and proportions* (Hoboken, NJ: Wiley-Interscience).

Fruhstorfer, H. (1984). Thermal sensibility changes during ischemic nerve block. *Pain* 20, 355-361.

Gandevia, S.C., and Phegan, C.M.L. (1999). Perceptual distortions of the human body image produced by local anaesthesia, pain and cutaneous stimulation. *The Journal of physiology* 514, 609-616.

Geber, C., Breimhorst, M., Burbach, B., Egenolf, C., Baier, B., Fechir, M., Koerber, J., Treede, R.-D., Vogt, T., and Birklein, F. (2013). Pain in chemotherapy-induced neuropathy—more than neuropathic? *Pain* 154, 2877-2887.

Geber, C., Klein, T., Azad, S., Birklein, F., Gierthmuhlen, J., Hugel, V., Lauchart, M., Nitzsche, D., Stengel, M., and Valet, M., et al. (2011). Test-retest and interobserver reliability of quantitative sensory testing according to the protocol of the German Research Network on Neuropathic Pain (DFNS): a multi-centre study. *Pain* 152, 548-556.

Geber, C., Scherens, A., Pfau, D., Nestler, N., Zenz, M., Tolle, T., Baron, R., Treede, R.-D., and Maier, C. (2009). Zertifizierungsrichtlinien für QST-Labore. *Schmerz (Berlin, Germany)* 23, 65-69.

Gierthmuhlen, J., Enax-Krumova, E.K., Attal, N., Bouhassira, D., Cruccu, G., Finnerup, N.B., Haanpaa, M., Hansson, P., Jensen, T.S., and Freynhagen, R., et al. (2015). Who is healthy? Aspects to consider when including healthy volunteers in QST—based studies—a consensus statement by the EUROPAIN and NEURO-PAIN consortia. *Pain* 156, 2203-2211.

Gustorff, B., Sycha, T., Lieba-Samal, D., Rolke, R., Treede, R.-D., and Magerl, W. (2013). The pattern and time course of somatosensory changes in the human UVB sunburn model reveal the presence of peripheral and central sensitization. *Pain* 154, 586-597.

Haanpaa, M., and Treede, R.-D. (2012). Capsaicin for neuropathic pain: linking traditional medicine and molecular biology. *European neurology* 68, 264-275.

Hansen, C., Hopf, H.C., and Treede, R.D. (1996). Paradoxical heat sensation in patients with multiple sclerosis. Evidence for a supraspinal integration of temperature sensation. *Brain : a journal of neurology* 119 (Pt 5), 1729-1736.

Hansson, P., Backonja, M., and Bouhassira, D. (2007). Usefulness and limitations of quantitative sensory testing: clinical and research application in neuropathic pain states. *Pain* 129, 256-259.

Hatem, S.M., Attal, N., Ducreux, D., Gautron, M., Parker, F., Plaghki, L., and Bouhassira, D. (2010). Clinical, functional and structural determinants of central pain in syringomyelia. *Brain : a journal of neurology* 133, 3409-3422.

Hayman, M., and Kam, P.C. (2008). Capsaicin. A review of its pharmacology and clinical applications. *Current Anaesthesia & Critical Care* 19, 338-343.

Henrich, F., Magerl, W., Klein, T., Greffrath, W., and Treede, R.-D. (2015). Capsaicin-sensitive C- and A-fibre nociceptors control long-term potentiation-like pain amplification in humans. *Brain : a journal of neurology* 138, 2505-2520.

Hertz, P., Bril, V., Orszag, A., Ahmed, A., Ng, E., Nwe, P., Ngo, M., and Perkins, B.A. (2011). Reproducibility of in vivo corneal confocal microscopy as a novel screening test for early diabetic sensorimotor polyneuropathy. *Diabetic Medicine* 28, 1253-1260.

Higgins, J.P.T., and Thompson, S.G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine* 21, 1539-1558.

Hirschfeld, G., Zernikow, B., Kraemer, N., Hechler, T., Aksu, F., Krumova, E., Maier, C., Magerl, W., and Blankenburg, M. (2012). Development of somatosensory perception in children: a longitudinal QST-study. *Neuropediatrics* 43, 10-16.

Hucho, T., and Levine, J.D. (2007). Signaling pathways in sensitization: toward a nociceptor cell biology. *Neuron* 55, 365-376.

Ichikawa, K., Koyama, N., Kiguchi, S., Kojima, M., and Yokota, T. (2001). Inhibitory effect of oxcarbazepine on high-frequency firing in peripheral nerve fibers. *European Journal of Pharmacology* 420, 119-122.

Isak, B., Pugdahl, K., Karlsson, P., Tankisi, H., Finnerup, N.B., Furtula, J., Johnsen, B., Sunde, N., Jakobsen, J., and Fuglsang-Frederiksen, A. (2017). Quantitative sensory testing and structural assessment of sensory nerve fibres in amyotrophic lateral sclerosis. *Journal of the neurological sciences* 373, 329-334.

Jørum, E., Warncke, T., Ziegler, E.A., Magerl, W., Fuchs, P.N., Meyer, R., and Treede, R.D. (2000). Secondary hyperalgesia to punctate stimuli is mediated by A-fiber nociceptors. In *Proceedings of the 9th World Congress on Pain, Progr Pain Res Management*, Devor M, Rowbotham M, Wiesenfeld-Hallin Z, ed. (Seattle: IASP Press), pp. 215–223.

Kandel, E.R., Schwartz, J.H., Jessel, T.M., Siegelbaum, S.A., Hudspeth, A.J. and Mack, S. (2013). *Principles of neural science* (New York, Lisbon, London: McGraw-Hill Medical).

- Katz, J., Finnerup, N.B., and Dworkin, R.H. (2008). Clinical trial outcome in neuropathic pain: relationship to study characteristics. *Neurology* 70, 263-272.
- Katz, N.P., Mou, J., Paillard, F.C., Turnbull, B., Trudeau, J., and Stoker, M. (2015). Predictors of Response in Patients With Postherpetic Neuralgia and HIV-Associated Neuropathy Treated With the 8% Capsaicin Patch (Qutenza). *The Clinical journal of pain* 31, 859-866.
- Kennedy, D.L., Kemp, H.I., Ridout, D., Yarnitsky, D., and Rice, A.S.C. (2016). Reliability of conditioned pain modulation: a systematic review. *Pain* 157, 2410-2419.
- Kilo, S., Schmelz, M., Koltzenburg, M., and Handwerker, H.O. (1994). Different patterns of hyperalgesia induced by experimental inflammation in human skin. *Brain : a journal of neurology* 117 (Pt 2), 385-396.
- Klein, T., Magerl, W., Rolke, R., and Treede, R.-D. (2005). Human surrogate models of neuropathic pain. *Pain* 115, 227-233.
- Koltzenburg, M., Torebjörk, H.E., and Wahren, L.K. (1994). Nociceptor modulated central sensitization causes mechanical hyperalgesia in acute chemogenic and chronic neuropathic pain. *Brain : a journal of neurology* 117 (Pt 3), 579-591.
- Koppert, W., Dern, S.K., Sittl, R., Albrecht, S., Schüttler, J., and Schmelz, M. (2001). A new model of electrically evoked pain and hyperalgesia in human skin: the effects of intravenous alfentanil, S(+)-ketamine, and lidocaine. *Anesthesiology* 95, 395-402.
- Krumova, E.K., Geber, C., Westermann, A., and Maier, C. (2012a). Neuropathic pain: is quantitative sensory testing helpful? *Current diabetes reports* 12, 393-402.
- Krumova, E.K., Zeller, M., Westermann, A., and Maier, C. (2012b). Lidocaine patch (5%) produces a selective, but incomplete block of Adelta and C fibers. *Pain* 153, 273-280.
- Lang, S., Klein, T., Magerl, W., and Treede, R.-D. (2007). Modality-specific sensory changes in humans after the induction of long-term potentiation (LTP) in cutaneous nociceptive pathways. *Pain* 128, 254-263.
- Lewis, G.N., Rice, D.A., and McNair, P.J. (2012). Conditioned pain modulation in populations with chronic pain: a systematic review and meta-analysis. *The journal of pain : official journal of the American Pain Society* 13, 936-944.

Loeser, J.D., and Treede, R.-D. (2008). The Kyoto protocol of IASP Basic Pain Terminology. *Pain* 137, 473-477.

Lotsch, J., Dimova, V., Hermens, H., Zimmermann, M., Geisslinger, G., Oertel, B.G., and Utsch, A. (2015). Pattern of neuropathic pain induced by topical capsaicin application in healthy subjects. *Pain* 156, 405-414.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In (The Regents of the University of California).

Magerl, W., Krumova, E.K., Baron, R., Tolle, T., Treede, R.-D., and Maier, C. (2010). Reference data for quantitative sensory testing (QST): refined stratification for age and a novel method for statistical comparison of group data. *Pain* 151, 598-605.

Magerl, W., and Treede, R.-D. (2004). Secondary tactile hypoesthesia: a novel type of pain-induced somatosensory plasticity in human subjects. *Festschrift dedicated to Prof. Manfred Zimmermann on the occasion of his 70th birthday* 361, 136-139.

Maier, C., Baron, R., Tolle, T.R., Binder, A., Birbaumer, N., Birklein, F., Gierthmuhlen, J., Flor, H., Geber, C., and Hugel, V., et al. (2010). Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): somatosensory abnormalities in 1236 patients with different neuropathic pain syndromes. *Pain* 150, 439-450.

Mainka, T., Bischoff, F.S., Baron, R., Krumova, E.K., Nicolas, V., Pennekamp, W., Treede, R.-D., Vollert, J., Westermann, A., and Maier, C. (2014). Comparison of muscle and joint pressure-pain thresholds in patients with complex regional pain syndrome and upper limb pain of other origin. *Pain* 155, 591-597.

Mainka, T., Malewicz, N.M., Baron, R., Enax-Krumova, E.K., Treede, R.-D., and Maier, C. (2016). Presence of hyperalgesia predicts analgesic efficacy of topically applied capsaicin 8% in patients with peripheral neuropathic pain. *European journal of pain (London, England)* 20, 116-129.

Marskey, H., Able Fessard, D.G., and Bonica, J.J. (1979). Pain terms; A list with definitions and notes on usage. *Pain* 6, 52.

Martinez, V., Fletcher, D., Martin, F., Orlikowski, D., Sharshar, T., Chauvin, M., Bouhassira, D., and Attal, N. (2010). Small fibre impairment predicts neuropathic pain in Guillain-Barre syndrome. *Pain* 151, 53-60.

McDonnell, A., Schulman, B., Ali, Z., Dib-Hajj, S.D., Brock, F., Cobain, S., Mainka, T., Vollert, J., Tarabar, S., and Waxman, S.G. (2016). Inherited erythromelalgia due to mutations in SCN9A: natural history, clinical phenotype and somatosensory profile. *Brain : a journal of neurology* 139, 1052-1065.

McLean, M.J., Schmutz, M., Wamil, A.W., Olpe, H.-R., Portet, C., and Feldmann, K.F. (1994). Oxcarbazepine. Mechanisms of Action. *Epilepsia* 35, S5-S9.

Mendell, L.M. (2014). Constructing and Deconstructing the Gate Theory of Pain. *Pain* 155, 210-216.

Ochoa, J.L. (1993). The human sensory unit and pain: new concepts, syndromes, and tests. *Muscle & nerve* 16, 1009-1016.

Ochoa, J.L., Campero, M., Serra, J., and Bostock, H. (2005). Hyperexcitable polymodal and insensitive nociceptors in painful human neuropathy. *Muscle & nerve* 32, 459-472.

Orstavik, K., Namer, B., Schmidt, R., Schmelz, M., Hilliges, M., Weidner, C., Carr, R.W., Handwerker, H., Jorum, E., and Torebjork, H.E. (2006). Abnormal function of C-fibers in patients with diabetic neuropathy. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 26, 11287-11294.

Pacaud, D., Romanchuk, K.G., Tavakoli, M., Gougeon, C., Virtanen, H., Ferdousi, M., Nettel-Aguirre, A., Mah, J.K., and Malik, R.A. (2015). The Reliability and Reproducibility of Corneal Confocal Microscopy in Children. *Investigative ophthalmology & visual science* 56, 5636-5640.

Pfau, D.B., Krumova, E.K., Treede, R.-D., Baron, R., Toelle, T., Birklein, F., Eich, W., Geber, C., Gerhardt, A., and Weiss, T., et al. (2014). Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): reference data for the trunk and application in patients with chronic postherpetic neuralgia. *Pain* 155, 1002-1015.

Rand, W.M. (1971). Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association* 66, 846-850.

Randic, M., Jiang, M.C., and Cerne, R. (1993). Long-term potentiation and long-term depression of primary afferent neurotransmission in the rat spinal cord. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 13, 5228-5241.

- Reimer, M., Hullemann, P., Hukauf, M., Keller, T., Binder, A., Gierthmuhlen, J., and Baron, R. (2017). Prediction of response to tapentadol in chronic low back pain. *European journal of pain (London, England)* 21, 322-333.
- Rolke, R., Andrews Campbell, K., Magerl, W., and Treede, R.-D. (2005). Deep pain thresholds in the distal limbs of healthy human subjects. *European journal of pain (London, England)* 9, 39-48.
- Rolke, R., Baron, R., Maier, C., Tolle, T.R., Treede, R.-D., Beyer, A., Binder, A., Birbaumer, N., Birklein, F., and Botefur, I.C., et al. (2006a). Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): standardized protocol and reference values. *Pain* 123, 231-243.
- Rolke, R., Magerl, W., Campbell, K.A., Schalber, C., Caspari, S., Birklein, F., and Treede, R.-D. (2006b). Quantitative sensory testing: a comprehensive protocol for clinical trials. *European journal of pain (London, England)* 10, 77-88.
- Rousseeuw, P.J. (1987). Silhouettes. A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20, 53-65.
- Schilder, A., Magerl, W., Hoheisel, U., Klein, T., and Treede, R.-D. (2016). Electrical high-frequency stimulation of the human thoracolumbar fascia evokes long-term potentiation-like pain amplification. *Pain* 157, 2309-2317.
- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics* 6, 461-464.
- Self, S.G., and Liang, K.-Y. (1987). Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions. *Journal of the American Statistical Association* 82, 605.
- Serra, J., Bostock, H., Sola, R., Aleu, J., Garcia, E., Cokic, B., Navarro, X., and Quiles, C. (2012). Microneurographic identification of spontaneous activity in C-nociceptors in neuropathic pain states in humans and rats. *Pain* 153, 42-55.
- Simone, D.A., Sorkin, L.S., Oh, U., Chung, J.M., Owens, C., LaMotte, R.H., and Willis, W.D. (1991). Neurogenic hyperalgesia: central neural correlates in responses of spinothalamic tract neurons. *Journal of neurophysiology* 66, 228-246.

Simpson, D.M., Schifitto, G., Clifford, D.B., Murphy, T.K., Durso-De Cruz, E., Glue, P., Whalen, E., Emir, B., Scott, G.N., and Freeman, R. (2010). Pregabalin for painful HIV neuropathy: a randomized, double-blind, placebo-controlled trial. *Neurology* 74, 413-420.

Smith, A.G., Howard, J.R., Kroll, R., Ramachandran, P., Hauer, P., Singleton, J.R., and McArthur, J. (2005). The reliability of skin biopsy with measurement of intraepidermal nerve fiber density. *Journal of the neurological sciences* 228, 65-69.

Smith, S.M., Dworkin, R.H., Turk, D.C., Baron, R., Polydefkis, M., Tracey, I., Borsook, D., Edwards, R.R., Harris, R.E., and Wager, T.D., et al. (2017). The Potential Role of Sensory Testing, Skin Biopsy, and Functional Brain Imaging as Biomarkers in Chronic Pain Clinical Trials: IMMPACT Considerations. *The journal of pain : official journal of the American Pain Society*.

Squire, L., Bloom, F.E., and Spitzer, N.C. (2008). *Fundamental Neuroscience* (s.l.: Elsevier professional).

Tesfaye, S., Boulton, A.J.M., and Dickenson, A.H. (2013). Mechanisms and management of diabetic painful distal symmetrical polyneuropathy. *Diabetes care* 36, 2456-2465.

Themistocleous, A.C., Ramirez, J.D., Shillo, P.R., Lees, J.G., Selvarajah, D., Orengo, C., Tesfaye, S., Rice, A.S.C., and Bennett, D.L.H. (2016). The Pain in Neuropathy Study (PiNS): a cross-sectional observational study determining the somatosensory phenotype of painful and painless diabetic neuropathy. *Pain* 157, 1132-1145.

Tramer, M.R., and Walder, B. (2005). Number needed to treat (or harm). *World journal of surgery* 29, 576-581.

Treede, R.D., Meyer, R.A., Raja, S.N., and Campbell, J.N. (1992). Peripheral and central mechanisms of cutaneous hyperalgesia. *Progress in neurobiology* 38, 397-421.

Treede, R.-D., Jensen, T.S., Campbell, J.N., Cruccu, G., Dostrovsky, J.O., Griffin, J.W., Hansson, P., Hughes, R., Nurmikko, T., and Serra, J. (2008). Neuropathic pain: redefinition and a grading system for clinical and research purposes. *Neurology* 70, 1630-1635.

Truini, A., Biasiotta, A., Di Stefano, G., La Cesa, S., Leone, C., Cartoni, C., Leonetti, F., Casato, M., Pergolini, M., and Petrucci, M.T., et al. (2013). Peripheral nociceptor

sensitization mediates allodynia in patients with distal symmetric polyneuropathy. *Journal of neurology* 260, 761-766.

Truini, A., Padua, L., Biasiotta, A., Caliandro, P., Pazzaglia, C., Galeotti, F., Inghilleri, M., and Cruccu, G. (2009). Differential involvement of A-delta and A-beta fibres in neuropathic pain related to carpal tunnel syndrome. *Pain* 145, 105-109.

Voets, T., Droogmans, G., Wissenbach, U., Janssens, A., Flockerzi, V., and Nilius, B. (2004). The principle of temperature-dependent gating in cold- and heat-sensitive TRP channels. *Nature* 430, 748-754.

Vollert, J., Attal, N., Baron, R., Freynhagen, R., Haanpaa, M., Hansson, P., Jensen, T.S., Rice, A.S.C., Segerdahl, M., and Serra, J., et al. (2016a). Quantitative sensory testing using DFNS protocol in Europe: an evaluation of heterogeneity across multiple centers in patients with peripheral neuropathic pain and healthy subjects. *Pain* 157, 750-758.

Vollert, J., Kramer, M., Barroso, A., Freynhagen, R., Haanpaa, M., Hansson, P., Jensen, T.S., Kuehler, B.M., Maier, C., and Mainka, T., et al. (2016b). Symptom profiles in the painDETECT Questionnaire in patients with peripheral neuropathic pain stratified according to sensory loss in quantitative sensory testing. *Pain* 157, 1810-1818.

Vollert, J., Maier, C., Attal, N., Bennett, D.L.H., Bouhassira, D., Enax-Krumova, E.K., Finnerup, N.B., Freynhagen, R., Gierthmühlen J, J., and Haanpää, M., et al. (2017a). Stratifying patients with peripheral neuropathic pain based on sensory profiles: algorithm and sample size recommendations. *Pain*, ePub ahead of print.

Vollert, J., Maier, C., Baron, R., Binder, A., Enax-Krumova, E.K., Geisslinger, G., Gierthmühlen, J., Hüllemann, P., Klein, T., and Lotsch, J., et al. (2017b). Sensory phenotype profiles in human surrogate models for denervation, peripheral, and central sensitization: a multi-center study. in preparation.

Vollert, J., Mainka, T., Baron, R., Enax-Krumova, E.K., Hüllemann, P., Maier, C., Pfau, D.B., Tolle, T., and Treede, R.-D. (2015). Quality assurance for Quantitative Sensory Testing laboratories: development and validation of an automated evaluation tool for the analysis of declared healthy samples. *Pain* 156, 2423-2430.

von Hehn, C.A., Baron, R., and Woolf, C.J. (2012). Deconstructing the neuropathic pain phenotype to reveal neural mechanisms. *Neuron* 73, 638-652.

- Wasner, G., Kleinert, A., Binder, A., Schattschneider, J., and Baron, R. (2005). Postherpetic neuralgia: topical lidocaine is effective in nociceptor-deprived skin. *Journal of neurology* 252, 677-686.
- Wasner, G., Schattschneider, J., Binder, A., and Baron, R. (2004). Topical menthol—a human model for cold pain by activation and sensitization of C nociceptors. *Brain : a journal of neurology* 127, 1159-1171.
- Westermann, A., Krumova, E.K., Pennekamp, W., Horch, C., Baron, R., and Maier, C. (2012). Different underlying pain mechanisms despite identical pain characteristics: a case report of a patient with spinal cord injury. *Pain* 153, 1537-1540.
- Woolf, C.J. (2011). Central sensitization: implications for the diagnosis and treatment of pain. *Pain* 152, S2-15.
- Woolf, C.J., Bennett, G.J., Doherty, M., Dubner, R., Kidd, B., Koltzenburg, M., Lipton, R., Loeser, J.D., Payne, R., and Torebjork, E. (1998). Towards a mechanism-based classification of pain? *Pain* 77, 227-229.
- Woolf, C.J., and Mannion, R.J. (1999). Neuropathic pain: aetiology, symptoms, mechanisms, and management. *Lancet (London, England)* 353, 1959-1964.
- Wopking, S., Scherens, A., Haussleiter, I.S., Richter, H., Schuning, J., Klauenberg, S., and Maier, C. (2009). Significant difference between three observers in the assessment of intraepidermal nerve fiber density in skin biopsy. *BMC neurology* 9, 13.
- Yarnitsky, D., and Granot, M. (2006). Chapter 27 Quantitative sensory testing. In *Pain*, F. Cervero and T.S. Jensen, eds. (Edinburgh, New York: Elsevier), pp. 397–409.
- Yarnitsky, D., and Ochoa, J.L. (1990). Release of cold-induced burning pain by block of cold-specific afferent input. *Brain : a journal of neurology* 113 (Pt 4), 893-902.
- Yarnitsky, D., and Ochoa, J.L. (1991). Differential effect of compression-ischaemia block on warm sensation and heat-induced pain. *Brain : a journal of neurology* 114 (Pt 2), 907-913.
- Youden, W.J. (1950). Index for rating diagnostic tests. *Cancer* 3, 32-35.

Ziegler, E.A. (1999). Secondary hyperalgesia to punctate mechanical stimuli. Central sensitization to A-fibre nociceptor input. *Brain* 122, 2245-2257.

Ziegler, E.A., Magerl, W., Meyer, R.A., and Treede, R.D. (1999). Secondary hyperalgesia to punctate mechanical stimuli. Central sensitization to A-fibre nociceptor input. *Brain : a journal of neurology* 122 (Pt 12), 2245-2257.

Zimmermann, M. (1968). Dorsal root potentials after C-fiber stimulation. *Science (New York, N.Y.)* 160, 896-898.

Zweig, M.H., and Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical chemistry* 39, 561-577.

6.1 Publications that arose from the master thesis prefacing this work

Vollert, J.*, Kramer, M.*, Barroso, A., Freynhagen, R., Haanpää, M., Hansson, P., Jensen, T.S., Kuehler, B.M., Maier, C., Mainka, T., Reimer, M., Segerdahl, M., Serra, J., Solà, R., Tölle, T.R., Treede, R.-D., and Baron, R. (2016b). Symptom profiles in the painDETECT Questionnaire in patients with peripheral neuropathic pain stratified according to sensory loss in quantitative sensory testing. *Pain* 157, 1810-1818.

Vollert, J., Mainka, T., Baron, R., Enax-Krumova, E.K., Hüllemann, P., Maier, C., Pfau, D.B., Tolle, T., and Treede, R.-D. (2015). Quality assurance for Quantitative Sensory Testing laboratories: development and validation of an automated evaluation tool for the analysis of declared healthy samples. *Pain* 156, 2423-2430.

6.2 Publications that arose from this work

Baron R.*, Maier C.*, Attal N., Binder A., Bouhassira D., Cruccu G., Finnerup N.B., Haanpää M., Hansson P., Hüllemann P., Jensen T.S., Freynhagen R., Kennedy J.D., Magerl W., Mainka T., Reimer M., Rice A.S.C., Segerdahl M., Serra J., Sindrup S.H., Sommer S., Tölle T.R., Vollert J.*, and Treede R.D.* (2017). Peripheral neuropathic pain: a mechanism-related organizing principle based on sensory profiles. *Pain* 158, 261-272.

Vollert, J., Attal, N., Baron, R., Freynhagen, R., Haanpää, M., Hansson, P., Jensen, T.S., Rice, A.S.C., Segerdahl, M., Serra, J., Sindrup, S.H., Tölle, T.R., Treede, R.-D., and Maier, C. (2016a). Quantitative sensory testing using DFNS protocol in Europe: an evaluation of heterogeneity across multiple centers in patients with peripheral neuropathic pain and healthy subjects. *Pain* 157, 750-758.

Vollert J.*, Maier C.*, Attal N., Bennett D.L.H., Bouhassira D., Enax-Krumova E.K., Finnerup N.B., Freynhagen R., Gierthmühlen J., Haanpää M., Hansson P., Hüllemann P., Jensen T.S., Magerl W., Ramirez J.D., Rice A.S.C., Schuh-Hofer S., Segerdahl M., Serra J., Shillo P.R., Sindrup S., Tesfaye S., Themistocleous A.C., Tölle T.R. TR, Treede R.-D.*, and Baron R.* (2017a). Stratifying patients with peripheral neuropathic pain based on sensory profiles. *Pain*, published ahead of print.

Vollert, J., Maier, C., Baron, R., Binder, A., Enax-Krumova, E.K., Geisslinger, G., Gierthmühlen, J., Hüllemann, P., Klein, T., Lötsch, J., Magerl, W., Oertel, B., Schuh-Hofer, S., Tölle, T.R., Treede, R.D. (2017b). Sensory phenotype profiles in human surrogate models for denervation, peripheral, and central sensitization: a multi-center study. in preparation.

*contributed equally.

7 LIST OF TABLES AND FIGURES

7.1 Tables

Table 1: Participants under human surrogate models of neuropathic pain included in the analysis.

Table 2: I^2 Index of heterogeneity, ranging from 0% (no heterogeneity) to 100% (perfect heterogeneity) for all eleven normally distributed QST parameters for healthy participants, polyneuropathy patients and patients with peripheral nerve injury.

Table 3: Decision on the number of clusters in patients.

Table 4: Patient cluster characteristics.

Table 5: Mean QST z-values and standard deviations for each of the 13 QST parameters separately for each of the three phenotypes.

Table 6: Crosstabulation of dominant phenotype identified using cluster analysis vs. the proposed new, individualized algorithm for full and simplified phenotyping.

Table 7: Crosstabulation of deterministic and probabilistic identification of thermal hyperalgesia phenotype in relation to irritable and non-irritable nociceptor phenotype in (Demant et al., 2014).

Table 8: Frequency of each phenotype in diabetic polyneuropathy, peripheral nerve injury and post-herpetic neuralgia, separately for the deterministic and probabilistic algorithm, and for full and simplified phenotyping.

Table 9: Number of patients, that need to be screened to find a sub-population with a given phenotype large enough to conduct a study with a power of 80% with an alpha-level of 0.05 and a given effect size.

Table 10: z-scores and effect sizes (D) of human surrogate models.

Table 11: Decision on the number of clusters for human surrogate models.

Table 12: Mean QST z-values and standard deviations for each of the 13 QST parameters separately for each of the three mechanistic subgroups.

Table 13: Crosstabulation of dominant phenotype identified using the algorithm vs. the original assignment for training and test dataset.

Table 14: Crosstabulation of dominant phenotype identified using the algorithm developed in patients vs. the algorithm developed based on surrogate models for n=902 patients.

7.2 Figures

Figure 1: CONSORT flowchart of data inclusion for the European cohort.

Figure 2: Forest plot of center-specific and overall mean QST z-values with 95% confidence intervals for healthy participants, patients with polyneuropathy and peripheral nerve injury.

Figure 3: Sensory profiles of the three-cluster solution for test and replication data sets.

Figure 4: Cluster separation projected onto two-dimensional space.

Figure 5: Frequencies of abnormal QST findings for the test dataset (n = 902).

Figure 6: Distribution of the three clusters within each neuropathic pain etiology.

Figure 7: ROC (Receiver Operating Characteristics) analysis the discriminatory power of the healthy probability to separate between patients with neuropathic pain and healthy participants.

Figure 8: Phenotype probabilities and probability of being healthy for n = 902 patients with neuropathic pain and n = 188 healthy participants.

Figure 9: Venn and bar diagrams of phenotype frequency and overlap between phenotypes for diabetic polyneuropathy, peripheral nerve injury and post-herpetic neuralgia.

Figure 10: Deep sensory profiling of human surrogate models of a priori mechanisms.

Figure 11: Deep sensory profiling of additional human surrogate models.

Figure 12: Average profiles per mechanism-stratified human surrogate models.

Figure 13: Deterministic and probabilistic sorting of human surrogate models and patients suffering from peripheral neuropathic pain.

Figure 14: Average profiles per mechanism-stratified patient group.

8 CURRICULUM VITAE

Full name: Jan Vollert
Date of birth: 11.12.1985
City of birth: Altena (Westf.), Germany
Civil status: Single
Father: Dietmar Vollert
Mother: Karin Vollert

SCHOOL CAREER

1996 – 2005 Albert-Schweitzer-Gymnasium Plettenberg
June 17th, 2005 **Abitur (1.9)**

UNIVERSITY CAREER

WS2007/2008 - Study of Molecular Biology,
SS2014 University of Applied Sciences, Gelsenkirchen
August 19th, 2010 **Bachelor of Science (1.6)**
Bachelor thesis: Development of an extendable software platform for eQTL studies and phenotype-association-analysis (1.0)
August 21st, 2014 **Master of Science (1.8)**
Master thesis: Development of a Europe-wide database on neuropathic pain - From information integration and data clearing to subset analyses (1.7)

8.1 Publications

2017

Corneal Confocal Microscopy to Evaluate Small Nerve Fibers in Patients with HIV. Kemp HI, Petropoulos IN, Rice ASC, **Vollert J**, Maier C, Sturm D, Schargus M, Peto T, Hau S, Chopra R, Malik RA. **JAMA Ophthalmol**, 2017, Epub ahead of print

Stratifying patients with peripheral neuropathic pain based on sensory profiles: algorithm and sample size recommendations. **Vollert J***, Maier C*, Attal N, Bennett DLH, Bouhassira D, Enax-Krumova EK, Finnerup NB, Freynhagen R, Gierthmühlen J, Haanpää M, Hansson P, Hüllemann P, Jensen TS, Magerl W, Ramirez JD, Rice ASC, Schuh-Hofer S, Segerdahl M, Serra J, Shillo PR, Sindrup S, Tesfaye S, Themistocleous AC, Tölle TR, Treede RD*, Baron R*. **PAIN**, 2017, Epub ahead of print

High test-retest-reliability of pain-related evoked potentials (PREP) in healthy subjects. Özgül ÖS, Maier C, Enax-Krumova EK, **Vollert J**, Fischer M, Tegenthoff M, Höffken O. **Neurosci Lett**, 2017, Epub ahead of print

Remote ischaemic conditioning decreases blood flow and improves oxygen extraction in patients with early complex regional pain syndrome. Hegelmaier T, Kumowski N, Mainka T, **Vollert J**, Goertz O, Lehnhardt M, Zahn PK, Maier C, Kolbenschlag J. **Eur J Pain**, 2017, Epub ahead of print

Recovery of mechanical detection thresholds after direct digital nerve repair versus conduit implantation. Huber JL, Homann HH, Mainka T, Mannil L, **Vollert J**, Maier C. **J Hand Surg Eur Vol**, 2017, Epub ahead of print

Unimpaired endogenous pain inhibition in the early phase of complex regional pain syndrome. Kumowski N, Hegelmaier T, Kolbenschlag J, Maier C, Mainka T, **Vollert J**, Enax-Krumova E. **Eur J Pain**, 2017 May;21(5):855-865

Peripheral Neuropathic Pain: A mechanism-related organizing principle based on sensory profiles. Baron R*, Maier C*, Attal N, Binder A, Bouhassira D, Cruccu G, Finnerup NB, Haanpää M, Hansson P, Hüllemann P, Jensen TS, Freynhagen R, Kennedy JD, Magerl W, Mainka T, Reimer M, Rice ASC, Segerdahl M, Serra J, Sindrup SH, Sommer S, Tölle TR, **Vollert J***, Treede RD*. **PAIN**, 2017 Feb;158(2):261-272

2016

Short-term test-retest-reliability of conditioned pain modulation using the cold-heat-pain method in healthy subjects and its correlation to parameters of standardized quantitative sensory testing. Gehling J, Mainka T, **Vollert J**, Pogatzki-Zahn EM, Maier C, Enax-Krumova EK. **BMC Neurology**, 2016 Aug 5;16:125

Symptom profiles in the painDETECT questionnaire in patients with peripheral neuropathic pain stratified according to sensory loss in quantitative sensory testing. **Vollert J***, Kramer M*, Barroso A, Freynhagen R, Haanpää M, Hansson P, Jensen TS, Kuehler BM, Maier C, Mainka T, Reimer M, Segerdahl M, Serra J, Solà R, Tölle TR, Treede RD, Baron R. **PAIN**, 2016 Aug;157(8):1810-8

Inherited Erythromelalgia due to mutations in SCN9A: Natural History, Clinical Phenotype and Somatosensory Profile. McDonnell A, Schulman B, Ali Z, Dib-Hajj SD, Brock F, Cobain S, Mainka T, **Vollert J**, Tarabar S, Waxman SG. **BRAIN**, 2016, Apr;139(4):1052-65

Quantitative Sensory Testing using DFNS protocol in Europe: an evaluation of heterogeneity across multiple centers in patients with peripheral neuropathic pain and healthy subjects **Vollert J**, Attal N, Baron R, Freynhagen R, Haanpää M, Hansson P, Jensen TS, Rice ASC, Segerdahl M, Serra J, Sindrup SH, Tölle TR, Treede RD, Maier C. **PAIN**, 2016 Mar;157(3):750-8

2015

Quality assurance for QST-laboratories: Development and validation of an automated evaluation tool for the analysis of declared healthy samples. **Vollert J**, Mainka T, Baron R, Enax-Krumova EK, Hüllemann P, Maier C, Pfau DB, Tölle T, Treede RD. **PAIN**, 2015 Dec;156(12):2423-30

Sleep-disordered breathing decreases after opioid withdrawal: Results of a prospective controlled trial. Schwarzer A, Aichinger-Hinterhofer M, Maier C, **Vollert J**, Walther JW. **PAIN**, 2015 Oct; 156(10):2167–2174

Pain relief with lidocaine 5% patch in localized peripheral neuropathic pain in relation to pain phenotype. Demant D, Lund K, Finnerup NB, **Vollert J**, Maier C, Segerdahl M, Jensen TS, Sindrup SH. **PAIN**, 2015 Oct; 156(10):2234–2244

Predictors of the placebo analgesia response in randomized controlled trials of chronic pain: A meta-analysis of the individual data from nine industrially sponsored trials. Vase L*, **Vollert J***, Finnerup NB, Miao X, Atkinson G, Marshall S, Nemeth R, Lange B, Liss C, Price D, Maier C, Jensen TS, Segerdahl M. **PAIN**, 2015 Sep; 156(9):1795-802

2014

The effect of oxcarbazepine in peripheral neuropathic pain depends on pain phenotype. Demant D, Lund K, **Vollert J**, Maier C, Segerdahl M, Finnerup NB, Jensen TS, Sindrup SH. **PAIN**, 2014 Nov; 155(11):2263-73

Comparison of muscle and joint pressure-pain thresholds in patients with complex regional pain syndrome and upper limb pain of other origin. Mainka T, Bischoff FS, Baron R, Krumova EK, Nicolas V, Pennekamp W, Treede RD, **Vollert J**, Westermann A, Maier C. **PAIN**, 2014 Mar; 155(3):591-7

8.2 Oral presentations

New developments in the understanding and treatment of painful diabetic neuropathy: Stratification based on cluster analysis of large cohorts. Lecture at the 10th Congress of the European Pain Federation (EFIC) 2017, Copenhagen, Denmark

Establishing the diagnosis of small fibre neuropathy - getting under the skin or looking deep into the eyes. Moderation of the debate at the 6th International Congress on Neuropathic Pain 2017, Gothenburg, Sweden

Vom Phänotyp zur effektiven Therapie mit QST. Lecture at the 20th Benjamin-Franklin Day of Pain 2015, Berlin, Germany

8.3 Poster presentations

2017

Stratifying patients with peripheral neuropathic pain based on sensory profiles: algorithm and phenotype frequency in painful diabetic polyneuropathy. **Vollert J**, Attal N, Bennett DLH, Bouhassira D, Freynhagen R, Haanpää M, Hansson P, Jensen TS, Maier C, Ramirez JD, Rice ASC, Segerdahl M, Serra J, Shillo PR, Sindrup S, Tesfaye S, Themistocleous AC, Tölle TR, Treede RD, Baron R. 6th International Congress on Neuropathic Pain 2017, Gothenburg, Sweden

2016

Proposal to use of biomarker methods to enable stratification of patient populations in clinical trials for neuropathic pain. **Vollert J**, Baron R, Corradini L, Magerl W, Maier C, Serra J, Treede RD, Whiles A, Segerdahl M. 16th IASP World Congress on Pain 2016, Yokohama, Japan

Does intake of opioids increase the risk of fracture in the elderly? **Vollert J**, Kaisler M, Kipping K, Seybold D, Schwarzer A, Maier C. 16th IASP World Congress on Pain 2016, Yokohama, Japan

Somatosensory phenotype of 1090 patients with lesion of the nervous system in relation to Paradoxical Heat Sensations. **Vollert J**, Attal N, Baron R, Bouhassira D, Finnerup NB, Freynhagen R, Haanpää M, Hansson P, Jensen TS, Maier C, Rice A, Segerdahl M, Serra J, Sindrup S, Tölle TR, Treede RD. 16th IASP World Congress on Pain 2016, Yokohama, Japan

2015

Fühlen Männer anders? Genderunterschiede im somatosensorischen Profil von PatientInnen mit peripheren neuropathischen Schmerzsyndromen. **Vollert J**, Attal N, Baron R, Bouhassira D, Freynhagen R, Haanpää M, Hansson P, Jensen TS, Rice ASC, Segerdahl M, Serra J, Sindrup SH, Tölle TR, Treede RD, Maier C. German Pain Congress 2015, Mannheim, Germany

There is no systematic heterogeneity between the QST assessment of pain clinics across Europe. **Vollert J**, Attal N, Baron R, Bouhassira D, Freynhagen R, Haanpää

M, Hansson P, Jensen TS, Rice A, Segerdahl M, Serra J, Sindrup S, Treede RD, Tölle TR, Maier C. 5th International Congress on Neuropathic Pain 2015, Nice, France

2014

Improving diagnostic value of QST by directed selection of individual parameters. Vollert J, Baron R, Krumova EK, Mainka T, Tölle TR, Treede RD, Maier C. 15th IASP World Congress on Pain 2014, Buenos Aires, Argentina

Predictors of placebo analgesic response – a preliminary meta-analysis of 9 randomized, double-blind pain studies. Vollert J, Segerdahl M, Miao X, Atkinson G, Marshall S, Nemeth R, Lange B, Raines S, Finnerup N, Maier C, Jensen T, Vase L. 15th IASP World Congress on Pain 2014, Buenos Aires, Argentina

2013

Erhöhung der diagnostischen Aussagekraft von QST durch gerichtete Parameterauswahl. Vollert J, Baron R, Krumova EK, Mainka T, Tölle TR, Treede RD, Maier C, German Pain Congress 2013. Hamburg, Germany

2012

Wenn gesund nicht gesund ist – Entwicklung eines statistischen Testsystems zur Qualitätssicherung bei Probandenmessungen im QST-Labor. Vollert J, Krumova EK, Maier C. German Pain Congress 2012, Mannheim, Germany

9 ACKNOWLEDGMENTS

Since 2011, I had the chance to work in fantastic international consortia, which gave me the opportunity to learn from some of the most distinguished researchers in the field of pain research. Not having a medical background and only holding a B.Sc., they have given me the warmest possible welcome to the field, have always taken the time to explain the sometimes most simple basics of medicine and clinical research to a bioinformatician, and have valued my input as equal from day one. Without this group, and of course without their work over the last decades, this thesis would not have been possible, and my career in pain research, that started out as a summer job, might have ended early, if they had not infected me with their enthusiasm.

Of these, Christoph Maier and Rolf-Detlef Treede are of course the first I owe thankfulness to. Together with Ralf Baron, they lead me through this project, helped me shaping my ideas, and, on many occasions, put more trust in me than I had in myself. It would have been an honor to have one of them as a supervisor, let alone all three of them. Beyond these defining figures, Zahid Ali, Nanna Brix Finnerup, Andrew Rice, Märta Segerdahl, Thomas Tölle and Lene Vase have made congresses and consortial meetings, and Aarhus, a home to me.

In times of personal crisis, I can always rely on my parents, Karin and Dietmar Vollert, and my brother Nils Vollert, and they have never ceased putting trust in me. Peak performance can only be reached with a safety net, and I am glad that my family always provided safety and stability.

Over the last ten years, Katharina Möller has developed her university and professional career in close proximity to mine, she counseled me on career decisions and vice versa, she flew with me to congresses around the world, pushed me into my first stay abroad, and reminded me of my own ambitions when I seemed to lose them. Without her input and her drive, I would be a different person living a different life, and many milestones I could not have reached without her.

All these people, and many unnamed others, have contributed significantly to this work, which I am deeply thankful for.