ORIGINAL ARTICLE

Prediction of response to tapentadol in chronic low back pain

M. Reimer[1,a], P. Hüllemann[1,a], M. Hukauf[2], T. Keller[2], A. Binder[1], J. Gierthmühlen[1], R. Baron[1]

1 Division of Neurological Pain Research and Therapy, Department of Neurology, University Hospital Schleswig-Holstein, Kiel, Germany
2 StatConsult GmbH, Magdeburg, Germany

Correspondence
Ralf Baron
E-mail: r.baron@neurologie.uni-kiel.de

[a]MR and PH contributed equally.

Abstract

Background: Many chronic low back pain (cLBP) patients do not satisfactorily respond to treatment. The knowledge of responders and non-responders before initiating treatment would improve decision making and reduce health care costs. The aims of this exploratory prediction study in cLBP patients treated with tapentadol were to identify predictors of treatment outcome based on baseline characteristics, to evaluate quality-of-life and functionality as alternative outcome parameters and to develop nomograms to calculate the individual probability of response.

Methods: In a retrospective analysis of an open-label phase 3b trial, 46 baseline characteristics were included into statistical prediction modelling. One hundred and twenty-one patients were followed up during the titration and treatment period and 67 patients were analysed who discontinued the trial.

Results: Demographic data were not relevant for response prediction. Nine baseline co-variables were robust: painDETECT score, intensity of burning and painful attacks, SF36 Health Survey score (MCS, PCS), EuroQol-5, Hospital Anxiety/Depression Scale. Gender had a minor influence. Alternative outcomes (quality-of-life, functionality) were more important for response prediction than conventional pain intensity measures. Neuropathic symptoms (high painDETECT score) had a positive predictive validity. Painful attacks and classical yellow flags (depression, anxiety) negatively influenced the treatment response. High depression scores, female gender and low burning predicted discontinuation during titration.

Conclusion: In this exploratory study, predictive baseline characteristics have been identified that can be used to calculate the individual probability of tapentadol response in cLBP. The small sample size in relation to the number of initial variables is a limitation of this approach.

Significance: Predictors for treatment response of tapentadol were identified in patients with chronic low back pain based on clinical pre-treatment characteristics that can guide personalized treatment. Quality-of-life and functionality were the most relevant outcomes for response prediction.

1. Introduction

Chronic low back pain (cLBP) has a lifetime prevalence of more than 70%, it severely impacts quality-of-life and is the most common cause of limited activity (Hoy et al., 2012). Patients often do not respond to pharmacological treatment or suffer from unacceptable side effects. Several factors might influence this inter-patient variation of treatment effects, e.g. differences in mechanisms of pain generation, genetic differences or the functional or psychological state of the patient (cf. also Martini et al., 2013). Therefore, it is important to identify characteristics of patients which render an individual more responsive to a specific treatment (Baron et al., 2012). This patient selection is particularly important for the treatment of non-malignant pain conditions with opioids since in many countries there is an ongoing discussion of opioid over- and abuse. Furthermore, in the process of "shared decision making", physicians and patients both actively participate in deciding on choices for therapeutic options. Adequate communication about risks and benefits is a prerequisite. Clinical prediction models provide the evidence-based input for shared decision making, by calculating estimates of the individual probabilities of risks and benefits.

We performed a prediction study in patients with cLBP. Pain perception is influenced by complex interactions of biological, psychological and social factors (Hagen et al., 2006). Furthermore, cLBP is frequently complicated by neuropathic components to the overall pain experience, e.g. by a co-existing painful radiculopathy. Neuropathic back pain patients suffer from distinct sensory symptoms, e.g. burning, paresthesias and hypersensitivity to thermal stimuli (Attal et al., 2011; Förster et al., 2013; Martel et al., 2015). The painDETECT questionnaire which is validated for the use in back pain patients captures the characteristic neuropathic symptoms and can estimate the intensity of each symptom quantitatively (Freynhagen et al., 2006).

One pharmacological strategy to address the mechanistic complexity of low back pain is to take a multi-mechanistic approach using two or more analgesic agents with different mechanisms of action to produce additive or even synergistic effects. Tapentadol has been developed to combine two mechanisms of action, μ-opioid receptor agonism (MOR) and noradrenaline reuptake inhibition (NRI) in a single molecule. It is effective in patients with cLBP (Lange et al., 2010). Because mechanisms of descending noradrenergic modulation seem to be of particular importance in chronic neuropathic pain (Dickenson, 2014; Goncalves et al., 2015), tapentadol may be well suited for the management of back pain with a neuropathic component (Baron et al., 2015, 2016).

In clinical trials, treatment efficacy is generally defined as improvement of the pain intensity score and relies on the patients' ratings of spontaneous pain as the primary endpoint. It is very likely, however, that outcome parameters like improvement of quality-of-life or increase in functional ability might be more important for the patients' overall well-being. Thus, the identification of predictors that impact on these alternative outcome parameters would improve pain management in clinical practice (Mehta et al., 2015).

The aims of the present exploratory study were
(1) to identify predictors of response to tapentadol treatment in cLBP based on baseline clinical characteristics,
(2) to evaluate quality-of-life and functionality as alternative outcome parameters in cLBP and
(3) to develop nomograms based on multivariable models to predict the outcome after tapentadol treatment for individual patients.

2. Methods

This is a retrospective analysis of response prediction in patients with severe cLBP. The open-label, phase 3b trial KF5503/44 that evaluated the effectiveness and tolerability of tapentadol hydrochloride in subjects with uncontrolled severe chronic nociceptive, mixed or neuropathic low back pain was conducted between 2009 and 2010 and was used for the analysis. Patients received tapentadol prolonged release (50–250 mg bid) during a 5-week titration and 7-week maintenance period. The average tapentadol dose after titration was 311 mg. For the primary endpoint (reduction in pain intensity), the trial was positive (for details see (Steigerwald et al., 2012)). The pharmaceutical manufacturer of tapentadol provided all original data of the trial and allowed us to access the data without restrictions. The company supported the analysis financially but had no influence in the process. The statisticians were reimbursed by us and had no connection with the company.

The study was conducted in accordance with applicable local laws, the principles of the Declaration of Helsinki, and Good Clinical Practice guidelines. All patients signed an informed consent document prior to enrolling in the study. The

protocol, patient information sheet, informed consent documents and amendments were reviewed by independent ethics committees.

2.1 Baseline co-variables

Clinical characteristics of the patients that were available at the baseline visit before tapentadol treatment was started (baseline co-variables, influencing factors) were analysed separately or in combination for prediction of treatment response to tapentadol. Besides demographic parameters (e.g. gender, BMI, medical history, vital signs and physical examination), we focused on patients' self-reported characteristics (e.g. psychosocial functioning) and patient-reported symptoms (e.g. sleep disruption, neuropathic pain symptoms). The following questionnaires were used to assess these parameters: pain intensity score (NRS-3), Sleep Evaluation Questionnaire (SQ), Short Form 36 Health Survey (SF 36), subject's satisfaction with treatment, EuroQol-Health State Today (EQ5D VAS), Hospital Anxiety and Depression Scale (HADS, including subscores HADS-A and HADS-D) and the painDETECT questionnaire (PDQ). The PDQ has seven separate questions addressing specific symptoms, i.e. burning, paresthesias, mechanical allodynia, painful attacks, thermal hypersensitivity, numbness and pressure-evoked pain. In total, 46 baseline co-variables were used for the prediction analysis (see Table 1).

2.2 Outcome variables

The pain-NRS-response was defined as the change of the spontaneous pain intensity score (NRS) from baseline to final evaluation visit after 12 weeks. In addition to this conventional most frequently used endpoint, we investigated two additional outcome parameters after 12 weeks: quality-of-life and functional ability. The QoL-response defined as change in quality-of-life parameters derived from the SF36 MCS subscore (QoL-MCS-response) as well as from the EQ-5D questionnaire (QoL-EQ-5D-response) and the functionality-response defined as change in parameters of bodily functioning derived from SF-36 (PCS subscore used to estimate physical function, Function-PCS-response) were analysed between baseline and endpoint (12 weeks). Furthermore, the overall drop-outs and drop-outs because of adverse events were examined in relation to possible influencing and predicting baseline factors. Depending on linear or binary data, the determined outcome variables were analysed in separate prediction models (see below).

Table 1 Baseline co-variables used for the prediction analysis.

	Variable	Sub-variable/ category	Unit	All patients ($N = 121$)
1	Age (years)		Mean/SD	58.7/10.9
2	Height		Mean/SD	166.8/10.3
3	Weight		Mean/SD	82.7/16.8
4	BMI		Mean/SD	29.8/5.7
5	Pulse rate (beats/min)		Mean/SD	73.6/9.3
6	Systolic blood pressure (mmHg)		Mean/SD	135.6/14.9
7	Diastolic blood pressure (mmHg)		Mean/SD	81.1/8.5
8	Respiratory rate (×/min)		Mean/SD	16.1/4.2
9	Gender	Male	N/%	77/63.6
		Female		44/36.4
10	Concomitant disease	Yes	N/%	96/79.3
		No		25/20.7
11	Surgical and medical procedures	Yes	N/%	69/57.0
		No		52/43.0
12	Neurological signs for radiculopathy at examination	Yes	N/%	30/24.8
		No		91/75.2
13	Musculoskeletal signs at examination	Yes	N/%	43/35.5
		No		78/64.5
14	Pain Intensity Score (NRS-3)		Mean/SD	7.3/1.0
15	PainDETECT Questionnaire Score		Mean/SD	15.5/7.6
16	PDQ-Items	BUR	Mean/SD	2.0/1.7
17		PRI	Mean/SD	2.1/1.7
18		ALD	Mean/SD	1.3/1.5
19		ATT	Mean/SD	2.7/1.5
20		TER	Mean/SD	1.4/1.5
21		NMB	Mean/SD	2.0/1.7
22		PRS	Mean/SD	2.4/1.5
23	PDQ (categories)	Unlikely	N/%	41/36.0
		Unclear		30/26.3
		Likely		43/37.7
24	PDQ–Cluster1	Yes	N/%	18/15.4
		No		99/84.6
25	PDQ–Cluster2	Yes	N/%	23/19.7
		No		94/80.3
26	PDQ–Cluster3	Yes	N/%	24/20.5
		No		93/79.5
27	PDQ–Cluster4	Yes	N/%	36/30.8
		No		81/69.2
28	PDQ–Cluster5	Yes	N/%	16/13.7
		No		101/86.3
29	Subject's satisfaction with treatment		Mean/SD	0.6/0.5
30	HADS-Depression Scale		Mean/SD	7.4/4.4
31	HADS-Anxiety Scale		Mean/SD	7.4/4.3
32	SQ-Sleep Quality Score		Mean/SD	2.8/0.8
33	SQ-Wake up		Mean/SD	2.8/1.8
34	SQ-Sleep last night		Mean/SD	362.1/100.4
35	SQ-Bedtime last night		Mean/SD	5.9/3.3
36	SF36 – Physical Functioning		Mean/SD	37.9/21.4
37	SF36 – Role Physical		Mean/SD	22.5/35.7
38	SF36 – Bodily Pain		Mean/SD	27.6/15.6
39	SF36 – General Health		Mean/SD	45.7/18.4
40	SF36 – Vitality		Mean/SD	38.9/20.6

Table 1 (Continued)

	Variable	Sub-variable/ category	Unit	All patients (N = 121)
41	SF36 – Social Functioning		Mean/SD	57.3/28.9
42	SF36 – Role Emotional		Mean/SD	46.8/46.4
43	SF36 – Mental Health		Mean/SD	55.8/20.5
44	SF36 PCS		Mean/SD	32.5/7.7
45	SF36 MCS		Mean/SD	42.0/12.3
46	EQ5D-VAS		Mean/SD	51.9/20.5

NRS-3, Pain Intensity Score in the last three days; SQ, Sleep Evaluation Questionnaire; SF 36, Short Form 36 Health Survey; EQ5D VAS, Euro-Qol-Health State Today; HADS, Hospital Anxiety and Depression Scale; PDQ, painDETECT questionnaire; PD-Q clusters, subgroups of patients based on different painDETECT symptom profiles according to Förster et al. (2013) and Baron et al. (2012).

Patients who had available baseline data and who started with a titration of the treatment and could be followed up during the titration and maintenance period of the trial.

2.3 Statistical analysis

Descriptive statistics: Continuously scaled data were presented with mean, standard deviation, min, max, median, quartile 1 and quartile 3. Categorically scaled data were presented with absolute and relative frequencies.

2.3.1 Prediction models

The statistical prediction analysis was performed according to the TRIPOD statement and to recently published guidelines (Hickey et al., 2015). Since retrospective data with predetermined sample size were used in this study, a formal power analysis was not applicable. However, according to Harrell and Frank (Harrell and Frank, 2001), prediction models should contain a number of variables less than 1/20 of the number of cases. Taking into account 121 cases, the respective models should not contain more than six variables, which were fulfilled in the models.

2.3.1.1 Selection of predictors. In a first step, we evaluated the influence of a single variable at baseline on outcome with bivariable models (including the baseline value of the respective outcome variable) in order to identify possible predictors. In dependence on the scale of the respective outcome variable (continuously scaled or quasi-continuously scaled on the one hand and binary scaled on the other hand), linear and logistic models were applied. All possible relationships were mapped for evaluation of the validity of consecutively applied models (linear relationship, monotonic relationship). In parallel, a factor analysis was performed to identify variables with a high potential for collinearity. Among strongly correlating variables, the most important (largest factor-loading) ones were selected for further analysis. As a result, a set of potential predictors was identified which was included in a multivariable regression (linear or logistic regression in dependence on outcome variable, step two). Three selection processes (forward, backward, Lasso [only applicable for linear regression]) were applied. Predictors that were selected consistently (at least weakly significant in two out of three selection processes and at least highly significant in one) were used for consecutive analyses.

2.3.1.2 Characterization of models and validation. After establishing the set of possible predictors, the following analyses were performed to characterize and validate the models and to prevent overfitting.

The adjusted coefficient of determination (R^2_{adj} indicating the part of variability of the outcome which is explained by the prediction; linear models) as well as the *c*-statistics (=area under ROC curve; logistic models) was used to characterize the models.

Models with low values for the respective parameter ($R^2_{adj} < 0.3$, *c*-stat < 0.6) were regarded as irrelevant. Note that *c*-statistic = 0.5 and $R^2_{adj} = 0$ refer to missing as well as *c*-statistic = 1 and $R^2_{adj} = 1$ to perfect prediction, respectively.

In parallel, the models were reinvestigated by including the individual co-variables, from the most influential to the fewest. The *F*-change-test (linear models) as well as the likelihood test (logistic models) was used to characterize the variables, which significantly improved the prediction.

The robustness of the models was tested by excluding aberrant ranges of values by checking whether R^2_{adj} and *c*-statistics would considerably decrease (model is not robust) or remain stable (model is robust); this was performed by visualizing the relationship of outcome parameters and results of the model function (predicted values). These parameters were also included in the estimation of the optimism of the model by applying a previously published procedure (Steyerberg, 2009):

The whole selection process was applied on resampled datasets (via bootstrapping), and the gained models were applied on the original data. The difference between the related values for R2 adj and

c-statistics, respectively, characterize the potential for overfitting (optimism). Consequently, the adjusted $R^2_{adj,corr}$ and c-statistic$_{corr}$ formed the values corrected by optimism.

Two relevant co-variables were detected, which depend on each other showing opposite effects (i.e. painDETECT and painful attacks). Therefore, an additional analysis was performed using a "modified" painDETECT score without "painful attacks" (see Results section; Table 2).

2.3.1.3 Presentation of results. The results are presented by the equation belonging to the respective model together with the R^2_{adj} and c-statistic, respectively, and the values corrected by optimism $R^2_{adj,corr}$ and c-statistic$_{corr}$. The model function can be used to calculate the magnitude of response by a calculator or a spreadsheet software. In addition, the parameters of model equations using standardized values for predictors are tabulated (see Table 2) allowing a direct comparison of the strength of influence of the predicting variables.

All analyses were performed using SAS 9.2 (SAS Institute Inc., Car, NY, USA). The nomograms were calculated with software R V. 3.01 (R Core Team,

2015. R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org).

2.3.2 Formulas and nomograms to calculate individual response

The prediction models with the best and most robust predictive values are presented as nomograms (Harrell and Frank, 2001) and equation, respectively.

A nomogram can directly be used to calculate the predicted response. The nomogram visualizes the influence of the different predictive variables on different horizontal lines. Depending on the influence of each predictor, the different lines have different lengths. The longer a horizontal line the stronger is the influence. The influence of each predictor is visualized by a number of points on the respective horizontal line. By adding the points associated with each predictor, the anticipated magnitude of response can be read on the response horizontal line on the bottom of the nomogram (example with three variables in Fig. 1). For the exact calculation of the magnitude of predicted response, the following formula is applicable:

Table 2 Valid prediction models.

Baseline variable	Completers (n = 121)			Discontinuers (n = 38)
	Outcome variable			
	Function-PCS	QoL-EQ5D	QoL-MCS	Discontinuation during titration
HADS-A		−0.17*		
HADS-D			−0.41***	1.33**
EQ-5D VAS	−0.15	−0.73***		
painDETECT-6	0.27**	0.20**	0.38***	
PDQ burning				0.80*
PDQ attacks			−0.20*	
SF36 MCS	0.31***		−0.71***	
SF36 PCS	−0.32***			
Sex female	−0.22**			1.45
Coefficient of determination; c-statistics	R^2_{adj} = 0.31, $R^2_{adj,corr}$ = 0.21	R^2_{adj} = 0.54, $R^2_{adj,corr}$ = 0.47	R^2_{adj} = 0.41, $R^2_{adj,corr}$ = 0.35	c-statistic = 0.68, c-statistic$_{corr}$ = 0.61

Parameters of statistical models using the standardized variables. This type of presentation allows the comparison of the influence of the baseline variables on the outcome variables – the larger the absolute value of the respective parameter the stronger the influence. A significant relationship is indicated by *p < 0.05, **p < 0.01, ***p < 0.001. For all outcome measures, the corresponding baseline value has the highest impact on prediction (e.g. baseline MCS predicts outcome MCS, baseline PCS predicts outcome PCS). The results are shown for the models with the modified 6-item painDETECT score (painDETECT-6).

For the linear models, the coefficient of determination (R^2) is shown. The coefficient of determination is a number that indicates how well data fit a statistical model. An R^2 of 1 indicates that the regression line perfectly fits the data, while an R^2 of 0 indicates that the line does not fit the data at all (see text). For the binary model, the c-statistic is the measure for the global fit of the model. The c-statistic is also referred as area under the curve (AUC), which is the area under the ROC curve. The receiver operating characteristic (ROC) is a plot which illustrates the overall diagnostic performance by plotting the outcome related sensitivity vs. specificity for each value of the predictive model (Fig. 6).

NRS-3, Pain Intensity Score in the last three days; SF36 MCS, Short Form 36 Health Survey, mental component summary scale; SF36 PCS, Short Form 36 Health Survey, physical component summary scale; EQ5D VAS, EuroQol-Health State Today; HADS-A/D, Hospital Anxiety and Depression Scale; PDQ, painDETECT questionnaire; painDETECT-6, modified painDETECT score using six items excluding the item "painful attacks".
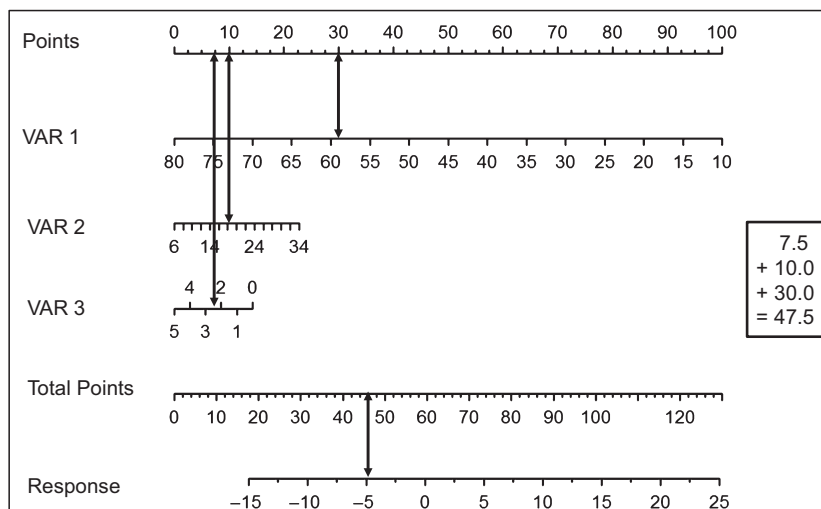
Figure 1 Example of a nomogram. A nomogram can directly be used to calculate the predicted response. The nomogram visualizes the influence of the different predictive variables on different horizontal lines. Depending on the influence of each predictor, the different lines have different lengths. The longer a horizontal line the stronger is the influence. The influence of each predictor is visualized by a number of points on the respective horizontal line. By adding the points associated with each predictor, the anticipated magnitude of response can be read on the response horizontal line on the bottom of the nomogram. For the calculation of a response, three variables are required in the example: Response = x * VAR1 + x * VAR2 + x * VAR3. Calculation example of this nomogram: VAR1 = 58 (30 Points), VAR2 = 18 (10 Points), VAR3 = 2.5 (7.5 Points), Total Points = 47.5 (30 + 10 + 7.5), Response = −5.

(1) Linear model (prediction of outcome):

$$\text{Response} = y = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n$$

(2) Binary model (prediction of discontinuation):

$$\text{Response} = y = \ln(P_x / 1 - P_x) = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n$$

The variables x_i could be differentially scaled (binary [0, 1], ordinal [handled as quasi continuous] and continuous.

3. Results

3.1 Patient cohorts and prediction models

All patients who had available baseline data and who started with a titration of the treatment ($N = 188$) were selected for further analysis. Outcome prediction data from 121 patients were analysed for establishing the models; these 121 patients were followed up during the titration and maintenance period of the trial (Table 1). In addition, 67 patients were analysed who dropped out of the study (38 patients discontinued during titration and 29 patients during the maintenance period; 36 out of 67 patients discontinued because of adverse events). The models were developed based on the entire study period (prediction of outcome) and on the titration period (prediction of discontinuation; see flowchart Fig. 2).

3.2 Baseline co-variables with predictive effect on outcome variables at endpoint

Significant associations could be identified between baseline co-variables and the following outcome variables assessing quality-of-life and functionality at the study endpoint:
(1) SF36 mental component summary scale (QoL-MCS-response)
(2) EQ-5D-VAS (QoL-EQ-5D-response)
(3) SF36 physical component summary scale (Function-PCS-response)

Interestingly, no significant and robust correlations could be found between any of the baseline co-variables and the outcome parameter "change in intensity of spontaneous pain" (pain-NRS-response).

After performing a multivariable analysis and the statistical correction procedures, nine out of the original 46 baseline co-variables still showed significant and robust associations with the outcome variables "quality-of-life and functionality" (Table 2):
(1) PainDETECT score (PDQ)
(2) PainDETECT sub-score: burning
(3) PainDETECT sub-score: attacks
(4) Short Form 36 Health Survey, mental component summary scale (SF36 MCS)
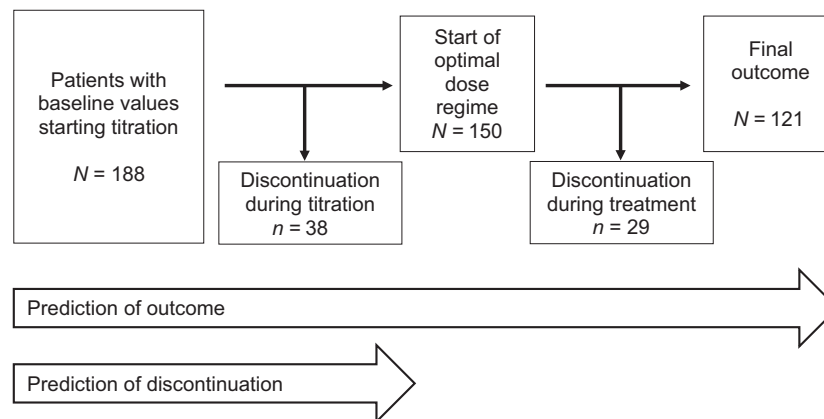(5) Short Form 36 Health Survey, physical component summary scale (SF36 PCS)

Figure 2 Flowchart of the study.

(6) EuroQol-5D-Health today overall (VAS)
(7) Hospital Anxiety and Depression Scale (HADS-A)
(8) Hospital Anxiety and Depression Scale (HADS-D)
(9) Gender

The painDETECT descriptor "painful attacks" had the opposite influence on prediction as compared with the other six painDETECT descriptors (burning, paresthesias, mechanical allodynia, thermal hypersensitivity, numbness and pressure-evoked pain). Therefore, a "modified" painDETECT score without "painful attacks" was additionally used in the models (PDQ-6 Score: PainDETECT score calculated with six items excluding painful attacks). In Table 2, the results are shown for the models with the modified 6-item PD-score.

3.3 Baseline co-variables with predictive effect on discontinuation of the treatment

For the outcome parameter "discontinuation" during the titration phase, three significant baseline co-variables with predictive potential could be identified:
(1) PainDETECT sub-score: burning
(2) Hospital Anxiety and Depression Scale (HADS-D)
(3) Gender

3.4 Detailed results of the prediction models

3.4.1 Prediction of improvement of functionality (Function-PCS-response)

In the entire study population, tapentadol improved the SF36 physical component summary scale (used to estimate physical function) significantly (+7.1 points). However, the improvement in functionality varied substantially between patients (minimum −13.8, maximum +29.5). Thus, some of the patients also deteriorate. Several baseline co-variables could

be identified with strong predictive effects on the improvement of physical function (Table 2). First, the corresponding baseline value (physical function at baseline) had the strongest impact, i.e. a low functionality value at baseline (low baseline PCS) predicted a high improvement in the same value (good PCS outcome). Furthermore, low baseline quality-of-life (EQ-5D VAS) was associated with a better outcome in functionality. Male patients had a slightly better functional response compared to women. Interestingly, there was no association between the baseline pain intensity and improvement of functionality. A high baseline painDETECT score (PDQ-6, high level of neuropathic pain) was associated with a high improvement of physical function (Table 2).

In order to calculate the estimated "Function-PCS-response" for individual patients, the weights of the different co-variables are described in the following formula:

$$\text{Function-PCS-response}$$
$$= 9.55 - 0.07 * \text{EQ-5D VAS} + 0.37 * \text{PDQ-6 Score}$$
$$- 0.23 * \text{SF36 MCS} - 0.37 * \text{SF36 PCS} - 4.23$$
$$* (\text{gender} = \text{female})$$

The nomogram which predicts the probability for improvement in functionality is shown in Fig. 3. It should be noted that the validity of this model is limited since the coefficient of determination decreases from 0.31 to 0.21 by applying the correction procedures.

3.4.2 Prediction of improvement of quality-of-life measured by EQ-5D VAS (QoL-EQ-5D-response)

In the entire study population, the EQ-5D VAS improved significantly (+16.8 points). The baseline EQ-5D VAS value had the strongest effect in the
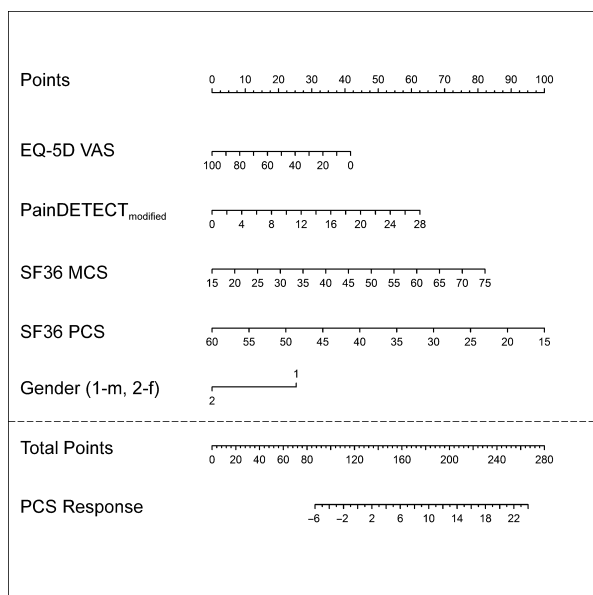
Figure 3 Nomogram of the Function-PCS-response. PainDETECT$_{modified}$: PainDETECT score using six items excluding the item "painful attacks" (painDETECT-6). m, male; f, female.

Figure 4 Nomogram of the QoL-EQ-5D-response. Details see Fig. 3.

prediction of an improvement of quality-of-life (Table 2). In addition, the following findings were demonstrated in the response formula and nomogram. The lower the baseline quality-of-life (EQ-5D VAS) and the anxiety level (HADS-A), the higher was the chance of an improvement of the QoL-EQ-5D-response. The painDETECT-6 score had a reverse effect on the response. The higher the score, i.e. the higher the likelihood for the presence of a neuropathic pain component, the higher was the chance of an improvement. The influence of all predictive parameters was visualized on a nomogram in Fig. 4. The formula for the prediction of the QoL-EQ-5D-response was calculated as follows:

$$\text{QoL-EQ-5D-response} = 68.25 - 1.10 * \text{HADS-A}$$
$$- 1.04 * \text{EQ-5D VAS}$$
$$+ 0.86 * \text{PDQ-6 Score}$$

3.4.3 Prediction of improvement of quality-of-life measured by SF-36 MCS (QoL-MCS-response)

In the entire study population, the SF-36 MCS improved significantly (+5.4 points). In addition to the corresponding baseline parameters, the co-variables HADS-D, PDQ-6 score and PDQ attacks had a significant influence on the QoL-MCS-response (Table 2). The lower the depression rate (HADS-D), intensity of painful attacks (PDQ attacks) and the mental component score (SF36 MCS) of patients, the
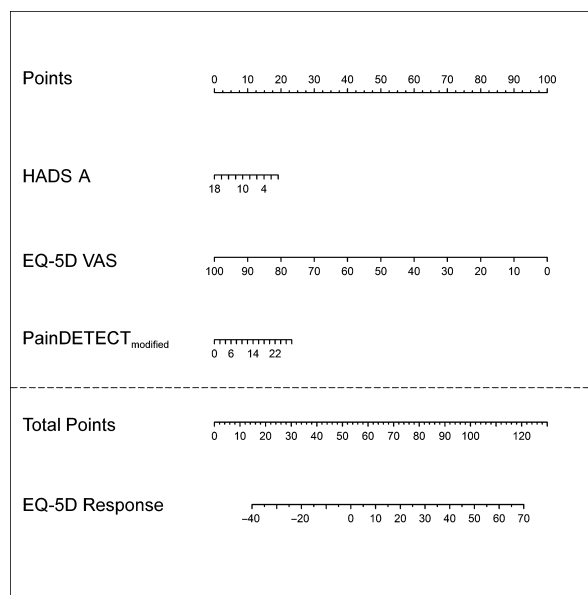
higher was the chance of an improvement of the QoL-MCS-response upon tapentadol treatment. The painDETECT-6 score had a reverse effect on the response, i.e. the higher the score, the higher was the chance of an improvement of the response. The corresponding nomogram is shown in Fig. 5. The formula for the prediction of the QoL-MCS-response was calculated as follows:

$$\text{QoL-MCS-response} = 35.39 - 1.02 * \text{HADS-D}$$
$$+ 0.62 * \text{PDQ-6 Score}$$
$$- 1.43 * \text{PD-Q attacks} - 0.63$$
$$* \text{SF36 MCS}$$

3.4.4 Prediction of discontinuation

The only valid prediction model for discontinuation was identified during the drug titration phase of the study. The presence of depression (HADS-D) had the strongest effect in this prediction model, followed by the intensity of a burning sensation (PDQ) and female gender (Table 2). These factors indicate that especially female patients and patients with high depression rates have a higher chance to discontinue. Furthermore, patients reporting severe burning sensations at the beginning have a higher chance to stay in the trial than other patients. The formula is the following:

Logit (Discontinuation of trial (yes))
$$= -1.84 + 0.12 * \text{HADS-D} - 0.25 * \text{PDQ burning}$$
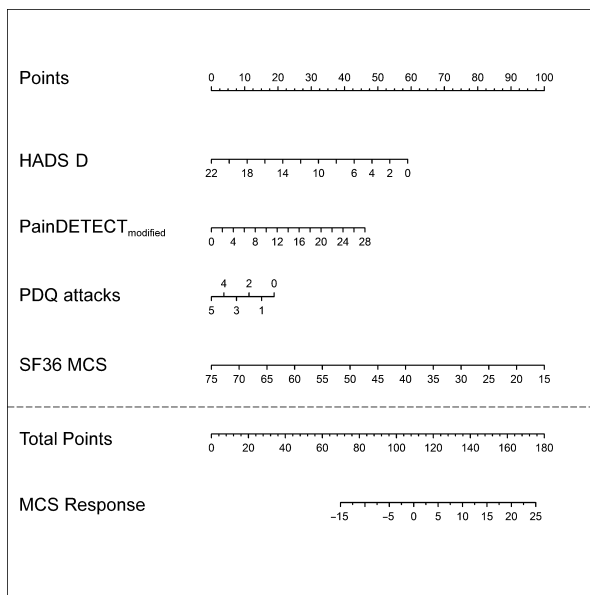$$+ 0.37 * (\text{gender} = \text{female})$$

Figure 5 Nomogram of the QoL-MCS-response. Details see Fig. 3.

The diagnostic accuracy of this prediction is illustrated in the receiver operating characteristics (ROC) curve (Fig. 6).

4. Discussion

The management of cLBP is still unsatisfactory despite a variety of pharmacological and non-pharmacological treatment options as well as multidisciplinary concepts of care (Artus et al., 2014). Thus, a considerable number of patients are treatment non-responders or discontinue the therapy because of adverse events. Therefore, treatments should be administered to those patients who will likely respond to them; response meaning demonstrating the most favourable risk-benefit ratio.

This exploratory study uses an extensive dataset of baseline characteristics derived from an open-label trial with tapentadol in cLBP to identify predictors of treatment response and of discontinuation. Furthermore, alternative outcome parameters, i.e. quality-of-life and functionality were tested in the prediction models. Because of the high amount of potential correlations, a very strict and conservative statistical modelling technique was applied. Six important findings can be summarized:

(1) Demographic baseline data were not relevant for response prediction.
(2) Gender had a minor influence on response prediction.
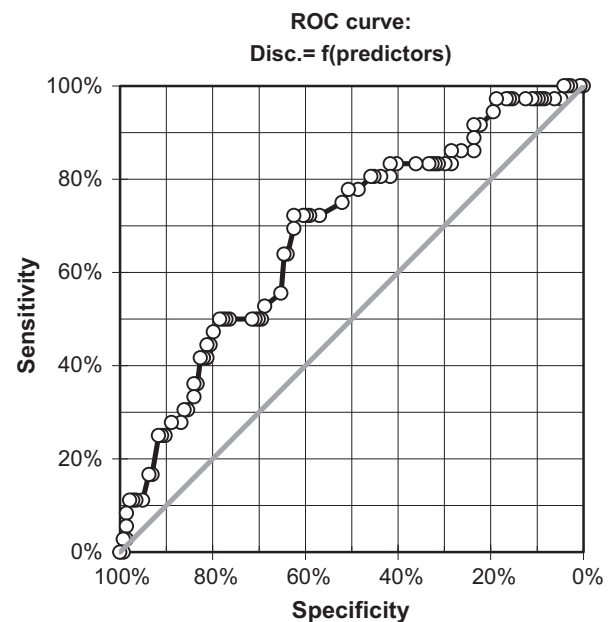(3) Alternative outcome parameters (quality-of-life, functionality) were more relevant for response

Figure 6 Receiver operating characteristic (ROC) curve for the model predicting discontinuation during titration period. A valid prediction model for discontinuation of the trial was identified during the drug titration period. Female patients and patients with high depression rates have a higher chance to discontinue. Patients with severe burning at the beginning have a higher chance to stay in the trial than other patients. The sensitivity and specificity values can be seen in the ROC curve.

prediction than conventional pain intensity measures.
(4) When alternative outcome parameters (quality-of-life, functionality) were used, the baseline painDETECT score had a positive and the intensity of painful attacks a negative predictive validity.
(5) Classical yellow flags, i.e. depression and anxiety, negatively influenced the treatment response.
(6) High depression scores, female gender and low burning pain predicted discontinuation during the titration phase.

4.1 Demographic characteristics and gender

Most of the demographic baseline characteristics did not influence the response to tapentadol in this trial. However, male patients had a slightly better chance to experience functional improvement under tapentadol treatment than females. Several studies have established gender differences in response to painful stimulation and in response to analgesic treatment (Hurley et al., 2008). For gender differences in opioid analgesia, multiple mechanisms including both pharmacokinetic and pharmacodynamic factors are

discussed (Craft, 2003). In accordance with our results, pharmacological increase of noradrenaline levels in the spinal cord of mice showed an analgesic effect only in male and not in female animals (Mifflin et al., 2016) pointing to a possible gender effect of the NRI action of tapentadol.

4.2 Alternative outcome parameters

If the usual, most commonly used outcome measure, i.e. reduction in spontaneous pain intensity, is used in the prediction models, no significant correlation could be demonstrated with any of the baseline co-variables. However, for the alternative outcome measures, quality-of-life and functionality, several robust predictors could be identified.

One reason for this unexpected finding might be the complex interplay of pain intensity on one hand and functionality and quality-of-life on the other hand. Once a treatment reduces the pain intensity, patients tend to be more active which in turn will increase the pain. Furthermore, it is conceivable that the reduction of spontaneous pain might not be the most important treatment goal for our patients. Many patients are willing to accept a certain amount of pain if they feel a substantial improvement in quality-of-life or functionality. Furthermore, the measure quality-of-life is a complex composite endpoint that combines an improvement of pain intensity and an improvement of functionality on one side and the negative impact of side effects on the other.

4.3 Baseline co-variable identical to outcome variable

The baseline measurement of an outcome parameter had the strongest predictive impact on the same outcome measure. For example, patients with a poor functionality at baseline had the best chance of improvement in functional performance. The finding is explained by the phenomenon of regression to the mean. This response parameter is regarded as a "general predictive effect", it is found in all clinical trials and is independent of a specific treatment (Dworkin et al., 2012). If future trials are designed with alternative primary outcome parameters like functionality or quality-of-life, a low baseline value should be used as an inclusion criterion.

4.4 Neuropathic symptoms and painful attacks

The second most robust predictor for a tapentadol response in functionality and quality-of-life was the total painDETECT score. In contrast to six PD-items (burning, paresthesias, mechanical allodynia, thermal hypersensitivity, numbness and pressure-evoked pain), the painDETECT descriptor "pain attacks" influences the outcome in a different direction. Therefore, we separated the information of attacks from the total painDETECT score and introduced a modified painDETECT-6 score without pain attacks.

In all models, patients with a high baseline painDETECT-6 sum score had a better chance to improve. These six painDETECT descriptors give some clues about the underlying pain mechanisms. If the painDETECT-6 score is high, i.e. a neuropathic component is likely, tapentadol has a better chance to improve the outcome. This is in-line with the mechanism of action of tapentadol and its additional noradrenaline re-uptake inhibition that has been demonstrated to be of particular benefit in neuropathic pain states (Hartrick and Rozek, 2011). Thus, the painDETECT-6 score can very likely be regarded as a "specific predictive effect" of tapentadol treatment.

Another robust but negative predictor is the existence of intense painful attacks measured with painDETECT. In several trials, tapentadol has been shown to decrease the intensity of attacks significantly (Steigerwald et al., 2012; Baron et al., 2016). However, the present results indicated that patients who suffer from many and severe attacks have a lesser chance to experience an improvement of QoL after tapentadol treatment. This result is in coincidence with the clinical experience that severe attacks are usually difficult to treat.

4.5 Psychological co-morbidities

In two prediction models, pre-treatment symptoms of anxiety or depression were identified as negative predictors for good outcome in quality-of-life. Furthermore, high depression rates predict drug discontinuation. These psychological factors belong to the classical "yellow flags" for chronicity of low back pain (Mallen et al., 2007). Several randomized controlled trials pointed to the same direction. Elevated pre-treatment HADS scores were associated with reduced opioid analgesic benefit and higher drop-out rates in trials with morphine and hydromorphone (Wasan et al., 2005, 2009, 2015; Jamison et al., 2013). One study supports the idea that the observed association is a "specific predictive effect" of opioid treatment since no such an effect was found in the placebo arm (Jamison et al., 2013). Two randomized controlled trials found low depression and anxiety scores ("high emotional functioning") as predictors

for a beneficial duloxetine response in painful diabetic neuropathy (Marchettini et al., 2016) and chemotherapy-induced peripheral neuropathy (Smith et al., 2015). Unfortunately, other important psychological variables with potential impact on outcome, like catastrophizing and kinesiophobia were not assessed in the present investigation.

4.6 Limitations of the study

Because of the limited sample size, an approach that separates patients into a subsample for developing the model and a second subsample for validation was not possible. Since the number of variables should be less than 1/20 of the sample size (Harrell and Frank, 2001), the full sample was necessary to develop the models. In order to address the problem of overfitting, we have prospectively planned to apply three approaches: exclusion of models with only minor measures of association, checking the models for robustness by excluding patients with very large or small values within a sensitivity analysis, calculating the optimism by comparing the model derived in original as well as on bootstrapped samples. Only models that successfully passed all three approaches were selected.

Since this analysis uses data of an open-label trial, a placebo-treated arm could not be incorporated into the study. Therefore, a clear distinction between general and drug-specific (treatment-dependent) predictive effects is not possible, a fact which is a clear limitation. However, for the use in daily clinical practice, it is necessary to identify all predictive effects that determine potential responders independently of being general or specific. Thus, the present approach closely mimics the approach of every physician who needs guidance in the selection of drugs that have a better likelihood to work for a specific patient.

5. Summary

Chronic back pain patients have a complex mixture of favourable and unfavourable predictors for treatment outcome. Response formulas and nomograms merge the available baseline parameters into one prediction model to calculate the individual probability of response. Despite the conservative statistical approach, some relevant predictors of tapentadol response were identified. Most importantly, we could demonstrate that alternative outcome parameters like quality-of-life and functionality might be more important for response prediction than

conventional pain intensity measures. Furthermore, the sensory neuropathic symptom profile that gives some insight into the underlying pain mechanisms had a strong predictive validity. Tapentadol, as anticipated from its mechanism of action, has a higher chance to produce pain reduction if more neuropathic symptoms are present. Furthermore, the classical yellow flags, psychological co-morbidities, negatively influence the treatment response.

In essence, this trial identifies predictors for tapentadol response and non-response based on clinical pre-treatment characteristics that might guide the clinical treatment decision in real life in the future.

Acknowledgements

References

Artus, M., van der Windt, D., Jordan, K.P., Croft, P.R. (2014). The clinical course of low back pain: a meta-analysis comparing outcomes in randomised clinical trials (RCTs) and observational studies. *BMC musculoskeletal disorders* 15: 68.

Attal, N., Perrot, S., Fermanian, J., Bouhassira, D. (2011). The neuropathic components of chronic low back pain: A prospective multicenter study using the DN4 Questionnaire. *J Pain* 12, 1080–1087.

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

Baron, R., Kern, U., Muller, M., Dubois, C., Falke, D., Steigerwald, I. (2015). Effectiveness and tolerability of a moderate dose of tapentadol prolonged release for managing severe, chronic low back pain with a neuropathic component: An Open-label Continuation Arm of a Randomized Phase 3b Study. *Pain Pract* 15, 471–486.

Baron, R., Likar, R., Martin-Mola, E., Blanco, F.J., Kennes, L., Muller, M., Falke, D., Steigerwald, I. (2016). Effectiveness of Tapentadol Prolonged Release (PR) compared with Oxycodone/Naloxone PR for the Management of Severe Chronic Low Back pain with a Neuropathic Component: A randomized, controlled, Open-Label, Phase 3b/4 Study. *Pain Pract* 16, 580–599.

Craft, R.M. (2003). Sex differences in drug- and non-drug-induced analgesia. *Life Sci* 72, 2675–2688.

Dickenson, A.H. (2014). Commentary on: Opioid and noradrenergic contributions of tapentadol in experimental neuropathic pain. *Neurosci Lett* 562, 90.

Dworkin, R.H., Turk, D.C., Peirce-Sandner, S., Burke, L.B., Farrar, J.T., Gilron, I., Jensen, M.P., Katz, N.P., Raja, S.N., Rappaport, B.A., Rowbotham, M.C., Backonja, M.M., Baron, R., Bellamy, N., Bhagwagar, Z., Costello, A., Cowan, P., Fang, W.C., Hertz, S., Jay, G.W., Junor, R., Kerns, R.D., Kerwin, R., Kopecky, E.A., Lissin, D., Malamut, R., Markman, J.D., McDermott, M.P., Munera, C., Porter, L., Rauschkolb, C., Rice, A.S., Sampaio, C., Skljarevski, V., Sommerville, K., Stacey, B.R., Steigerwald, I., Tobias, J., Trentacosti, A.M., Wasan, A.D., Wells, G.A., Williams, J., Witter, J., Ziegler, D. (2012). Considerations for improving assay sensitivity in chronic pain clinical trials: IMMPACT recommendations. *Pain* 153, 1148–1158.

Förster, M., Mahn, F., Gockel, U., Brosz, M., Freynhagen, R., Tolle, T.R., Baron, R. (2013). Axial low back pain: One painful area–many perceptions and mechanisms. *PLoS One* 8, e68273.

Freynhagen, R., Baron, R., Gockel, U., Tolle, T.R. (2006). painDETECT: A new screening questionnaire to identify neuropathic components in patients with back pain. *Curr Med Res Opin* 22, 1911–1920.

Goncalves, L., Friend, L.V., Dickenson, A.H. (2015). The influence of mu-opioid and noradrenaline reuptake inhibition in the modulation of pain responsive neurones in the central amygdala by tapentadol in rats with neuropathy. *Eur J Pharmacol* 749, 151–160.

Hagen, E.M., Svensen, E., Eriksen, H.R., Ihlebaek, C.M., Ursin, H. (2006) Comorbid subjective health complaints in low back pain. *Spine (Phila Pa 1976)* 31, 1491–1495.

Harrell, J., Frank, E. (2001). *Regression Modeling Strategies With Applications to Linear Models, Logistic Regression, and Survival Analysis.* (New York: Springer).

Hartrick, C.T., Rozek, R.J. (2011). Tapentadol in pain management: A mu-opioid receptor agonist and noradrenaline reuptake inhibitor. *CNS Drugs* 25, 359–370.

Hickey, G.L., Dunning, J., Seifert, B., Sodeck, G., Carr, M.J., Burger, H.U., Beyersdorf, F., Ejcts & Committees, I.E. (2015) Statistical and data reporting guidelines for the European Journal of Cardio-Thoracic Surgery and the Interactive CardioVascular and Thoracic Surgery. *Eur J Cardiothorac Surg* 48, 180–193.

Hoy, D., Bain, C., Williams, G., March, L., Brooks, P., Blyth, F., Woolf, A., Vos, T., Buchbinder, R. (2012). A systematic review of the global prevalence of low back pain. *Arthritis Rheum* 64, 2028–2037.

Hurley, R.W., Lesley, M.R., Adams, M.C., Brummett, C.M., Wu, C.L. (2008). Pregabalin as a treatment for painful diabetic peripheral neuropathy: A meta-analysis. *Reg Anesth Pain Med* 33, 389–394.

Jamison, R.N., Edwards, R.R., Liu, X., Ross, E.L., Michna, E., Warnick, M., Wasan, A.D. (2013). Relationship of negative affect and outcome of an opioid therapy trial among low back pain patients. *Pain Pract* 13, 173–181.

Lange, B., Kuperwasser, B., Okamoto, A., Steup, A., Haufel, T., Ashworth, J., Etropolski, M. (2010). Efficacy and safety of tapentadol prolonged release for chronic osteoarthritis pain and low back pain. *Adv Ther* 27, 381–399.

Mallen, C.D., Peat, G., Thomas, E., Dunn, K.M., Croft, P.R. (2007). Prognostic factors for musculoskeletal pain in primary care: A systematic review. *Br J Gen Pract* 57, 655–661.

Marchettini, P., Wilhelm, S., Petto, H., Tesfaye, S., Tolle, T., Bouhassira, D., Freynhagen, R., Cruccu, G., Lledo, A., Choy, E., Kosek, E., Mico, J.A., Spath, M., Skljarevski, V., Lenox-Smith, A., Perrot, S. (2016). Are there different predictors of analgesic response between antidepressants and anticonvulsants in painful diabetic neuropathy? *Eur J Pain* 20, 472–482.

Martel, M.O., Finan, P.H., Dolman, A.J., Subramanian, S., Edwards, R.R., Wasan, A.D., Jamison, R.N. (2015). Self-reports of medication side effects and pain-related activity interference in patients with chronic pain: A longitudinal cohort study. *Pain* 156, 1092–1100.

Martini, C.H., Yassen, A., Krebs-Brown, A., Passier, P., Stoker, M., Olofsen, E., Dahan, A. (2013). A novel approach to identify responder subgroups and predictors of response to low- and high-dose capsaicin patches in postherpetic neuralgia. *Eur J Pain* 17, 1491–1501.

Mehta, P., Claydon, L., Hendrick, P., Winser, S., Baxter, G.D. (2015). Outcome measures in randomized-controlled trials of neuropathic pain conditions: A systematic review of systematic reviews and recommendations for practice. *Clin J Pain* 31, 169–176.

Mifflin, K.A., Benson, C., Thorburn, K.C., Baker, G.B., Kerr, B.J. (2016). Manipulation of neurotransmitter levels has differential effects on formalin-evoked nociceptive behavior in male and female mice. *J Pain* 217, 483–498.

Steigerwald, I., Muller, M., Davies, A., Samper, D., Sabatowski, R., Baron, R., Rozenberg, S., Szczepanska-Szerej, A., Gatti, A., Kress, H.G. (2012). Effectiveness and safety of tapentadol prolonged release for severe, chronic low back pain with or without a neuropathic pain component: Results of an open-label, phase 3b study. *Curr Med Res Opin* 28, 911–936.

Steyerberg, E.W. (2009). *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating.* (New York, London: Springer).

Wasan, A.D., Davar, G., Jamison, R. (2005). The association between negative affect and opioid analgesia in patients with discogenic low back pain. *Pain* 117, 450–461.

Wasan, A.D., Butler, S.F., Budman, S.H., Fernandez, K., Weiss, R.D., Greenfield, S.F., Jamison, R.N. (2009). Does report of craving opioid medication predict aberrant drug behavior among chronic pain patients? *Clin J Pain* 25, 193–198.

Wasan, A.D., Michna, E., Edwards, R.R., Katz, J.N., Nedeljkovic, S.S., Dolman, A.J., Janfaza, D., Isaac, Z., Jamison, R.N. (2015). Psychiatric comorbidity is associated prospectively with diminished opioid analgesia and increased opioid misuse in patients with chronic low back pain. *Anesthesiology* 123, 861–872.