BJA

British Journal of Anaesthesia, 121 (5): 1123–1132 (2018)

doi: 10.1016/j.bja.2018.06.007
Advance Access Publication Date: 31 July 2018
Pain

Machine-learned selection of psychological questionnaire items relevant to the development of persistent pain after breast cancer surgery

J. Lötsch[1,2,*], R. Sipilä[3], V. Dimova[4] and E. Kalso[3]

[1]Institute of Clinical Pharmacology, Goethe University Frankfurt, Frankfurt am Main, Germany, [2]Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Branch for Translational Medicine and Pharmacology TMP, Frankfurt am Main, Germany, [3]Pain Clinic, Department of Anaesthesiology, Intensive Care and Pain Medicine, Helsinki University Hospital and University of Helsinki, Helsinki, Finland and [4]Department of Neurology, University Medical Centre of the Johannes Gutenberg University, Mainz, Germany

*Corresponding author. E-mail: j.loetsch@em.uni-frankfurt.de

Abstract

Background: Prevention of persistent pain after breast cancer surgery, via early identification of patients at high risk, is a clinical need. Psychological factors are among the most consistently proposed predictive parameters for the development of persistent pain. However, repeated use of long psychological questionnaires in this context may be exhaustive for a patient and inconvenient in everyday clinical practice.
Methods: Supervised machine learning was used to create a short form of questionnaires that would provide the same predictive performance of pain persistence as the full questionnaires in a cohort of 1000 women followed up for 3 yr after breast cancer surgery. Machine-learned predictors were first trained with the full-item set of Beck's Depression Inventory (BDI), Spielberger's State–Trait Anxiety Inventory (STAI), and the State–Trait Anger Expression Inventory (STAXI-2). Subsequently, features were selected from the questionnaires to create predictors having a reduced set of items.
Results: A combined seven-item set of 10% of the original psychological questions from STAI and BDI, provided the same predictive performance parameters as the full questionnaires for the development of persistent postsurgical pain. The seven-item version offers a shorter and at least as accurate identification of women in whom pain persistence is unlikely (almost 95% negative predictive value).
Conclusions: Using a data-driven machine-learning approach, a short list of seven items from BDI and STAI is proposed as a basis for a predictive tool for the persistence of pain after breast cancer surgery.

Keywords: persisting pain; psychological questionnaires; breast cancer; patients; machine-learning; data science

Editorial decision: 25 April 2018; **Accepted:** 25 April 2018

Editor's key points

- Identifying patients at risk of persistent pain after breast surgery would have clinical benefits.
- Validated questionnaires may be useful tools, but may create a significant burden for the patient.
- Machine learning was used to select key questions from several validated psychological questionnaires.
- A short questionnaire was generated, with performance assessed using an existing breast cancer surgery data set.
- A strength of this questionnaire is in identifying patients at lower risk of persistent pain.

Psychological factors play an important role in pain persistence[1–10] exerting bidirectional influences, that is, they influence how the patient perceives and interprets pain whereas constant pain may have an impact on mood either directly or via its negative effects on sleep, functionality, social, and other activities.[11,12] Thus, psychological factors may be at the same time predictors, maintainers, changeable variables, and consequences of persistent pain[13] and qualify as early predictive markers for the development of persisting pain.[14–18] However, the use of long questionnaires may be exhaustive for a patient and inconvenient in everyday clinical practice. Thus, there is need for a simplified yet accurate questionnaire which can also be used in a digital application. This has previously been addressed with regard to Spielberger's State–Trait Anxiety Inventory (STAI); however, this was carried out without focus on pain.[19] The goal of the present analysis was to create a simple questionnaire with a good predictive power for persisting pain after surgery. To achieve this, we used a new approach, machine learning, to reanalyse data from a 3-yr follow-up study that had used Beck's Depression Inventory (BDI), STAI, and the subscale Anger Inhibition from the State–Trait Anger Expression Inventory's (STAXI-2)[20] for the association with persistent pain in a cohort of 1000 women who had undergone breast cancer surgery.[10,21]

Methods

Subjects and study design

The study followed the Declaration of Helsinki and was approved by the Coordinating Ethics Committee of the Helsinki University Hospital (136/E0/2006). Each participant provided written informed consent. We enrolled women who had unilateral non-metastasised breast cancer treated at the Helsinki University Hospital between 2006 and 2010 with either breast-conserving surgery or mastectomy with axillary surgery (sentinel biopsy, axillary clearance, or both). Exclusion criteria were neoadjuvant therapy[22] and immediate breast reconstruction surgery. Of the 1536 consecutive eligible patients, 1149 patients were invited to participate, of whom 126 patients declined and 23 were withdrawn. The whole study cohort and the protocol have previously been described in detail.[10,21]

Assessment of post-surgical pain scores

Pain intensity was assessed at months 1, 6, 12, 24, and 36 after surgery using the Numerical Rating Scale (NRS) ranging from 0 (no pain) to 10 (the most severe pain that can be imagined).[23] Pain ratings were recorded using questionnaires sent to the patients and asking identical assessments of presence and intensity of pain in the areas of the previous breast cancer surgery (breast, axilla, and upper arm). The pain ratings acquired at 6 months or later after surgery were the basis for the patient's classification into the 'persistent pain' or the 'non-persistent pain' group. A period of 6 months was considered to more adequately reflect persistent post-surgery pain (PPSP) in the present clinical setting[17] than the original definition of PPSP proposing a shorter period of 2 months,[24] which seems premature for the diagnosis of chronic pain after breast cancer surgery as adjuvant therapies continue longer. Specifically, as used previously[25] patients with NRS ≤ 3 at 36 months after surgery ($NRS_{month36} \leq 3$) were identified as belonging to the 'non-persistent pain' group, whereas those with NRS > 3 at 36 months after surgery ($NRS_{month36} > 3$) belonged to the 'persistent pain' group. Further criteria for belonging to the 'non-persistent pain' group were the presence of no more than mild pain (i.e. $NRS_{month12..month36} \leq 3$), whereas the 'persistent pain' group was more precisely characterised by always having at least moderate pain without a consistent tendency to ameliorate [i.e. $NRS_{month36} > 3$ *and* $NRS_{month12..month36} > 0$ and $(NRS_{month36} - NRS_{month24}) \geq 0$].

Assessment of psychological parameters

Psychological factors (depressive symptoms, state and trait anxiety, and anger inhibition) were assessed with the respective standardised and validated questionnaires: BDI,[26] Spielberger's STAI,[27] and STAXI-2.[20] The time point was set to 6 months, as an earlier study (unpublished) identified this as the best time point for predicting the later outcome, hence providing the best start point for the present analysis aiming at reducing the number of questionnaire items without reducing their strength of association with persistent pain (or its absence).

Data analysis

Data were analysed using the R software package (version 3.4.1 for Linux; http://CRAN.R-project.org/)[28] on an Intel Xeon® computer running on Ubuntu Linux 16.04.3 64-bit. The analysis was performed in five main steps: (i) data preprocessing, (ii) feature selection, (iii) classifier creation from each set of questionnaire items followed by (iv) performance testing, and finally (v) the creation of a combined classifier from the identified item subsets of the questionnaires. This will be described briefly in the following discussion; detailed descriptions are provided in the Supplementary materials.

Data analysis used supervised machine learning,[29,30] which aims at learning a mapping from inputs x to outputs y, given a labelled set of input–output pairs. The inputs, 'features', consisted of the psychological questionnaire items (i.e. the 22, 21, 21, and 9 questions queried using the BDI, STAI-State, STAI-Trait, and STAXI-2 questionnaires, respectively). The outputs, 'classes', consisted of the two patient subgroups, those with or without persistent pain at 3 yr. The aim was to create a machine-learned classifier that uses a smaller number of questionnaire items and can still predict pain persistence, and the complete questionnaire.

In supervised machine learning, an algorithm is trained on data for which the class labels (persistent pain or not) of the patients are known, in order to correctly predict the class membership from data for which the class labels are

unknown. To this end, the data set was split into a training subset (two-thirds of the patients) and test subset (one-third of the patients), both containing patients with or without persistent pain in proportion to the size of the subset. Firstly, the algorithm was provided with the questionnaire items and the class information of the training data subset to learn to assign the class membership from the psychological features. Subsequently, the trained algorithm was used on the test data subset; however, it was only provided with the questionnaire items whereas the class information was omitted. The task was to assign the patients to the correct class (pain persistence group). However, as the class assignment was known for the test data subset, the performance of the trained machine-learned classifier could be quantified. Test sensitivity and specificity were calculated using standard equations[31] [i.e. sensitivity = true positives/(true positives + false negatives) and specificity = true negatives/(true negatives + false positives)]. In addition, the negative predictive value (NPV) indicating the likelihood that the patient is unlikely to develop persistent pain when the test is negative, was calculated as NPV = true negative/(true negative + false negative),[32] and the balanced test accuracy[33] was calculated as 0.5 × (true positive/all positive + true negative/all negative). To prevent the results from depending on a single random split of the data set, the analyses were performed 1000 times on the data subsets (training, test) randomly drawn in every run from the original data set by means of core-class proportional bootstrap[34] resampling.

The machine-learned classification algorithm was implemented as random forest analysis, which uses a multitude of decision trees to learn a highly irregular combination of features.[35,36] Feature selection was based on the importance of the psychological questionnaire items in the random forest classifier, which was obtained as the mean decrease in classification accuracy when the respective feature was excluded from forest building. The features included in the final classifier were identified using an item-categorisation technique implemented as computed ABC analysis,[37] which aims at dividing a set of positive data into three disjoint subsets called 'A', 'B', and 'C'. Subset 'A' comprises the profitable values (i.e. 'the important few'). The final size of the feature set was equal to the most frequent size of set 'A' in the 1000 runs. The final members of the feature set were chosen in decreasing order of their appearances in ABC set 'A' among the 1000 runs.

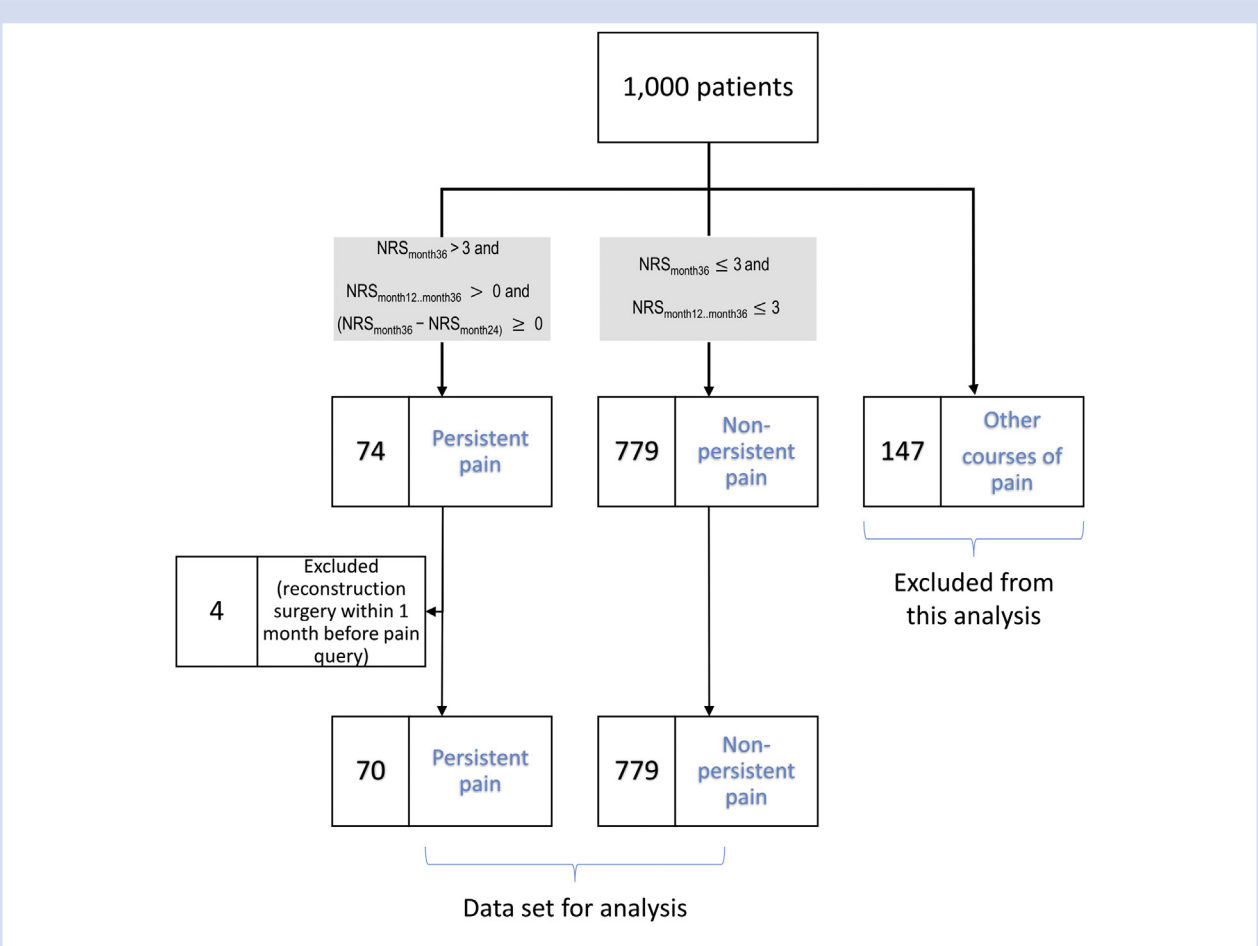

Fig 1. Flowchart showing the classification of the patients on the basis of the 3-yr development of pain after breast cancer surgery. A total of 853 women fell into the two main groups of persisting or non-persisting pain, according to the criteria displayed in the grey frames. This was the main cohort that was analysed. The remaining 143 women in whom the criteria for class assignment applied only partly were therefore excluded from machine-learned classifier establishment but they were used as an exploratory shortened 'test' data set.

These analyses were performed on each psychological questionnaire separately and subsequently, and the items so far identified as best suited for the correct group assignment of a subject were used to create a combined classifier. They were submitted to feature selection and classifier creation in the same manner as described above. To obtain similarly scaled items, STAI-State and STAI-Trait were rescaled into $n \in N,\ [0, \ldots, 3]$ by subtracting a value of 1 from each rating.

Results

The recovery rate of the pain questionnaires was high, being 95.3%, 91.3%, 90.2%, 90.2%, and 87.4% in months 1, 6, 12, 24, and 36, respectively. A flowchart of the subjects' inclusion in the data analysis is shown in Fig. 1. Persistent pain after breast cancer surgery was observed in 70 patients while the time course of post-surgery pain of 779 patients corresponded to the predefined criteria of non-persistent pain whereas class assignment criteria were not completely met in the remaining 143 women who were therefore excluded. Questionnaires completed by at least 80% of the items were required for subject inclusion into the data analysis. This allowed analysing 761, 752, 754, and 830 patients (for BDI, STAI-State, STAI-Trait, and STAXI-2, Anger Inhibition, respectively).

Performance of full and reduced questioners to identify patients with persistent pain

BDI

During the 1000 random forest analyses of bootstrap resampled data, different contributions to the overall classification of the patients to either the 'persistent pain' or the 'non-persistent pain' group were observed for particular items of the BDI queried at 6 months after breast cancer surgery (Fig. 2, left panel). In subsequent computed ABC analyses (Fig. 2, right panel), ABC set 'A' comprising items that contribute most to the separation of the two pain persistence groups, took sizes of d=3, 4, 5, or 6 BDI items in 1, 45, 427, and 527 of the runs, respectively. Therefore, a set size of d=6 BDI items was chosen for the final classifier (Fig. 3, top left panel). During further 1000 runs with resampled training data, all possible sums of the six BDI items, ranging from 0 to 18 (possible ratings of $n \in N,\ [0, \ldots, 3]$ per item), were tested as classifiers. The best result was found at item sums of less than 1, 2, 3, 4, or 5 occurring 74, 877, 33, 11, or 5 times, respectively. On this basis, the final classification rule was defined as if $\sum_1^6$BDI items < 2 then a subject belongs to the 'non-persistent pain' group else to the 'persistent pain' group (Fig. 3, right-hand side of the top left panel). The reduced BDI

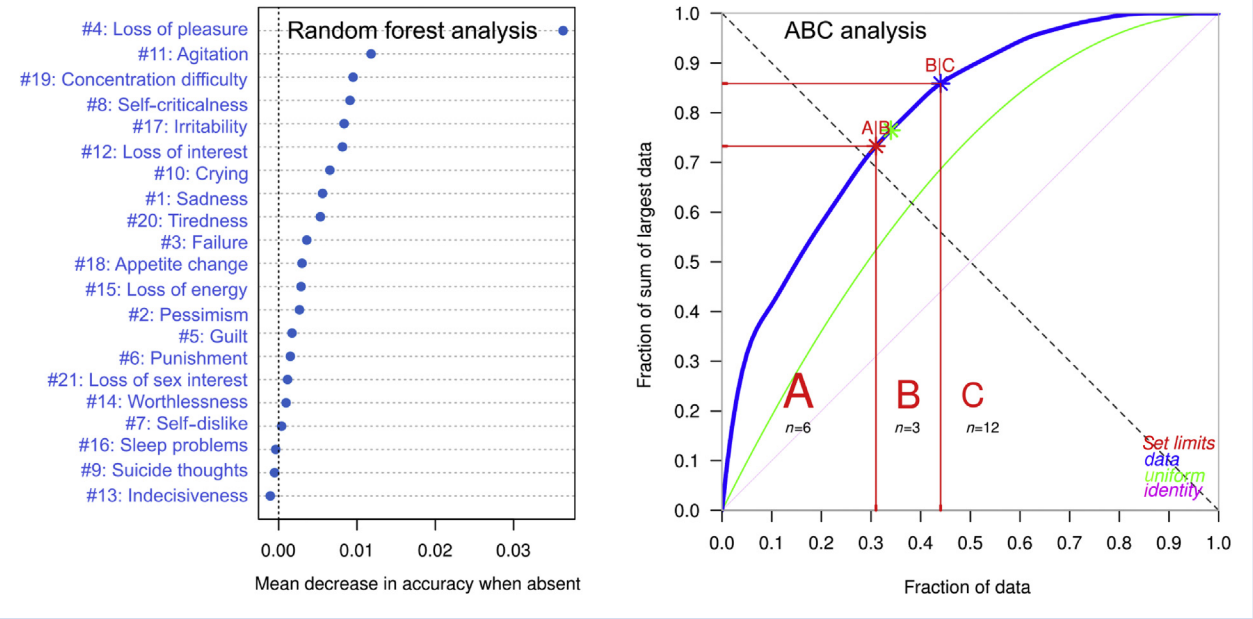

Fig 2. Random forest analysis followed by computed ABC analysis of the contribution of the questionaries' items to overall classification of the patients into either the 'persistent pain' or the 'non-persistent pain' group. To illustrate the approach, an example of the analysis of Beck's depression inventory is shown. Left panel: display of the mean decrease in classification accuracy when the respective feature (questionaries' item) is excluded from the random forest analysis. A mean decrease of zero is indicated with a dotted line. The plot displays one typical example out of the analyses of 1000 bootstrap resampled data subsets. Right panel: subsequent to random forest-based feature ranking, the mean decrease of accuracy associated with each item was submitted to computed ABC analysis, which is an item selection procedure aiming at identification of most profitable items from a larger list of items. The ABC plot (blue line) shows the cumulative distribution function of the mean decreases in accuracy, along with the identity distribution, x_i=constant (magenta line), that is each feature contributes similarly to the classification accuracy (for further details about computed ABC analysis, see Ultsch and Lötsch).[37] The red lines indicate the borders between ABC sets 'A', 'B', and 'C'. Only set 'A' containing the most profitable items was selected as the pain-relevant questionnaire subsets. The figure has been created using the R software package (version 3.4.1 for Linux; http://CRAN.R-project.org/).[28] In particular, the computed ABC analysis was performed and plotted using our R package 'ABCanalysis' (http://cran.r-project.org/package=ABCanalysis).[37]

item set provided a balanced classification accuracy of more than 60% to assign a patient to the correct group in the test data subset (Table 1). It was associated with a high NPV of the classifier of approximately 95%.

Spielberger's STAI, subscale STAI-State

For Spielberger's STAI subscale STAI-State, the mode of the number of items in ABC set 'A' was, with 827 out of 1000 runs with resampled data, observed at an item count of d=7 (Fig. 3). The classifier constructed from these seven items during further 1000 runs with resampled training data testing all possible sums of the items was defined as if $\sum_1^7 \text{STAI items} < 14$ then a subject belongs to the 'non-persistent pain' group else to the 'persistent pain' group (Fig. 3). This classifier provided similar balanced classification accuracy as the complete STAI-State questionnaire subscale (Table 1).

Spielberger's STAI, subscale STAI-Trait

For the subscale STAI-Trait, the mode of the number of items in ABC set 'A' was, with 861 out of 1000 runs with resampled data, observed at an item count of d=6 (Fig. 3, right top panel). During further 1000 runs with resampled training data, all possible sums of the six STAI-Trait items, ranging from 6 to 24 (possible ratings of $n \in N,\ [1, \ldots, 4]$ per item), the final classification rule was obtained as if $\sum_1^6 \text{STAITrait items} < 12$ then a subject belongs to the 'non-persistent pain' group else to the 'persistent pain' group. Again, its classification performance was similar to that of the complete STAI-Trait questionnaire (Table 1).

Spielberger's STAXI-2, subscale anger suppression (STAXI-2 Anger In)

For the STAXI-2 subscale Anger Inhibition, the mode of the number of items in ABC set 'A' was, with 861 out of 1000 runs

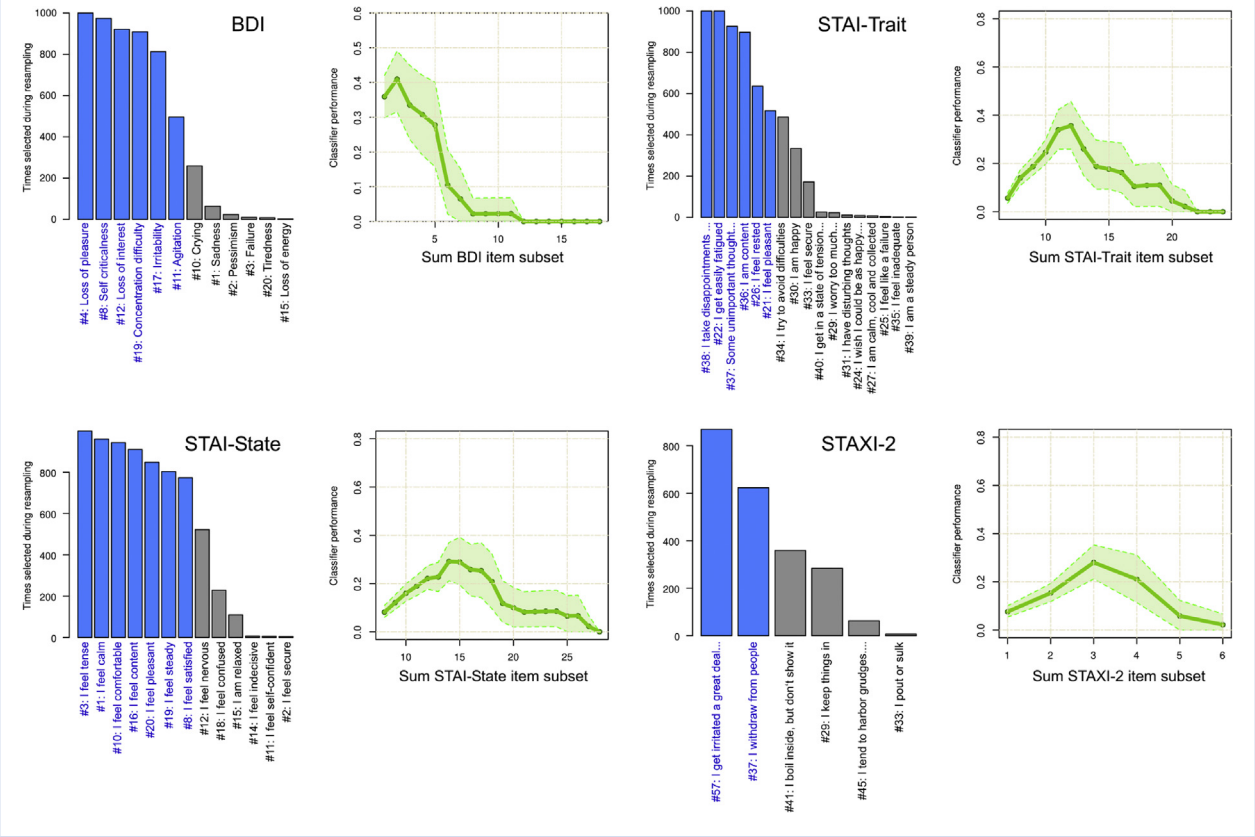

Fig 3. Items subsets of the psychological questionnaires and establishment of the classification rules for patient's assignment to either the 'persistent pain' or the 'non-persistent pain' group. The bar plots indicate how often the respective feature (questionnaire item) was found in the ABC set 'A' during 1000 random forest analyses with subsequent computed ABC analysis (Fig. 1) applied to 1000 data subsets randomly drawn from the original data sets by means of core-class proportional bootstrap resampling. The blue bars indicate those items that were selected for the creation of the reduced psychological questionnaire. The line plots at the right of each bar plot show the performance of different iterations of the questionnaires items subsets for the prediction of persistent pain after breast cancer surgery. All possible sums of the selected items, which can take values of $n \in N,\ [0, \ldots, 3]$ for BDI (shown) and STAXI-2, and $n \in N,\ [1, \ldots, 4]$ (not shown) for STAI–State and STAI-Trait (not shown), were iteratively tested with respect to their classification performance. The main the product of sensitivity and specificity. The lines show the product of sensitivity and specificity for different item sums. Given are the medians of 1000 bootstrap resampling runs, surrounded by the 95% bootstrap confidence intervals (2.5th to 97.5th percentiles). The classification rule tested was 'if the sum of the items < x', then the patient belongs to the 'non-persistent pain' group, else to the 'persistent pain' group. The analysis was done iteratively with item sums increasing by a value of 1 between each iteration. The figure was created using the R software package (version 3.4.1 for Linux; http://CRAN.R-project.org/).[28]

Table 1 Comparative test performance measures of reduced and complete psychological questionnaires for the prediction of persistent pain in a 3-yr follow-up after breast cancer surgery. The psychological parameters had been queried 6 months after surgery. The item reduction and creation of reduced questionnaires was performed on two-thirds of the data obtained via class-proportional random split of the original data set ('training data') whereas the assessment of predictive performance was performed with the one-third of the remaining data ('test data'). PPV, positive predictive value; NPV, negative predictive value; balanced accuracy = 0.5×(sensitivity + specificity); BDI, Beck's Depression Inventory; STAI, Spielberger's State–Trait Anxiety Inventory; STAXI-2, Spielberger's State–Trait Anger Expression Inventory

Test performance measure	Reduced questionnaire	Full questionnaire
BDI		
Sensitivity	0.61905	0.42857
Specificity	0.59130	0.70000
PPV	0.12150	0.11538
NPV	0.94444	0.93064
Balanced accuracy	0.60518	0.56429
STAI-A		
Sensitivity	0.71429	0.47619
Specificity	0.74890	0.59031
PPV	0.13889	0.14925
NPV	0.95714	0.93923
Balanced accuracy	0.65230	0.61254
STAI-B		
Sensitivity	0.52381	0.61905
Specificity	0.62719	0.64912
PPV	0.1145	0.13978
NPV	0.93464	0.94872
Balanced accuracy	0.57550	0.63409
STAXI-2		
Sensitivity	0.78261	0.60870
Specificity	0.42629	0.56175
PPV	0.11111	0.11290
NPV	0.95536	0.94000
Balanced accuracy	0.60445	0.58522

with resampled data, observed at an item count of d=2 (Fig. 3, right bottom panel). The classification rule derived from further 1000 runs with resampled training data testing all possible sums of the two STAXI-2 items, ranging from 0 to 6 (possible ratings of $n \in N$, $[0, \ldots, 3]$ per item) was defined as if $\sum_1^2$STAXI2 items < 3 then a subject belongs to the 'non-persistent pain' group else to the 'persistent pain' group (Table 1).

Combined reduced questionnaire

All items identified as best suited for pain persistence group association in each questionnaire were included as new candidate features in the combined score (Fig. 4). Again, feature selection was performed using random forests followed by computed ABC analysis, which resulted in a mode of the number of items in ABC set 'A', with 495 out of 1000 runs with resampled data, observed at an item count of d=7 (Table 2). In the combined score, the items most frequently assigned to ABC set 'A' comprised BDI item 8 ('self-criticalness'), STAI-State item 20 ('I feel pleasant'), item 10 ('I feel comfortable'), item 1 ('I feel calm'), STAI-Trait item 22 ('I get easily fatigued'), item 26 ('I feel rested'), and item 21 ('I feel pleasant'). The classification rule derived from further 1000 runs with resampled training data testing all possible sums of the seven items, ranging from 0 to 21 (possible ratings of $n \in N$, $[0, \ldots, 3]$ per item) was defined as if $\sum_1^7$items < 7 then a subject belongs to the 'non-persistent pain' group else to the 'persistent pain' group. The classification performance was similar to that of the reduced or complete questionnaires analysed above (Table 3).

Discussion

While preserving the predictive performance, the reduction of the questionnaires to a pain-relevant subset was considerable. The results suggest that only 25–35% of the items of the questionnaires are needed without compromising the performance of the trained algorithm for persistent pain. Moreover, the seven-item combination questionnaire derived from the best performing items of each questionnaire provided a comparable predictive performance to the original 69 items of the included questionnaires.

Of note, the presently obtained reduced sets of questions tended to exceed the predictive performance of the full questionnaires (Table 1). This is plausible when considering that all of the used questionnaires have been created to provide sum scores. The total sum score was higher in the patient group with persistent pain observed within the 3-yr follow-up after surgery than in patients with non-persistent pain. The majority of the questionnaire items, but not all, contributed to the higher total sum scores. For certain items the relation was reversed; some items were rated comparatively lower by the patients with persistent pain than by those in whom pain had taken a more favourable course. This is shown as an example for the BDI questionnaire in Supplementary Fig. S1. In a sum score, the inversely directed differences reduced the global group difference.

The present analyses produced a plausible list of psychological items agreeing with psychological parameters known to associate with pain. Specifically, the final combined seven-item subset includes a symptom cluster from two major psychological factors, depressive mood and both state and trait anxiety. These psychological characteristics can be categorised as: (i) heightened self-criticalness; (ii) anxious feelings of unpleasantness, uncomfortableness and nervousness during the previous week; and (iii) general tendency to feel being in an unrelaxed, fatigued, and unpleasant state. All of them represent components of the fear avoidance model of chronic pain.[38,39] The included items on state and trait anxiety cover also the same categories of characteristics of nervousness and dissatisfaction. Such attentional focusing on bodily experiences of anxiety may also activate and interact with hypervigilance to pain-related bodily signals leading to an overestimation of pain sensations and restarting the circle of rumination.

This cluster of anxiety-related characteristics can also be interpreted within the concept of anxiety sensitivity.[40,41] Anxiety sensitivity is known to attract attention to bodily experiences,[42] and therefore pain sensations may be more intense and distracting in an anxious person. Habituation to rumination may be associated with a less resilient way to cope with pain sensations[43,44] and possibly with higher catastrophising which is known to have an effect on pain experience.[45] Feelings of dissatisfaction and self-blaming may also influence the way how a person copes with the primary disease and pain. Tendency to hear only negative bodily

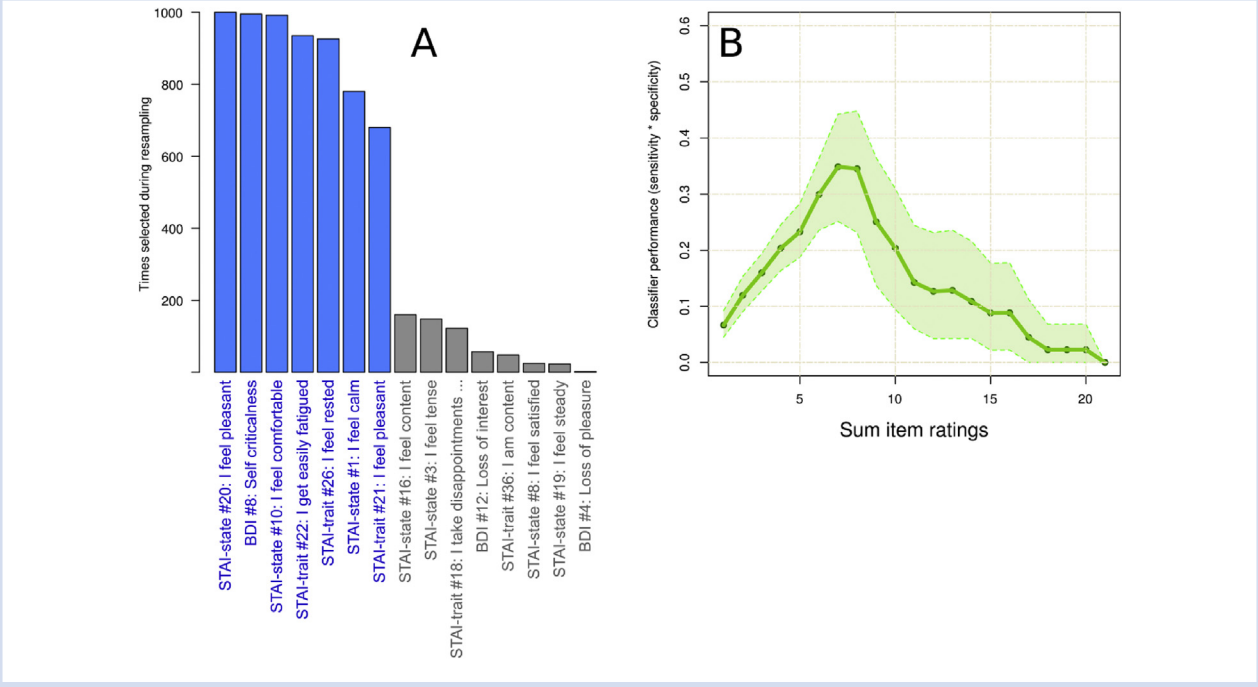

Fig 4. Creation a of combined seven-item questionnaire from the items selected in the previously analytical steps from each questionnaire as being best suited for the prediction of persistent pain. All items from each separate questionnaire that had been found most suitable for predicting persistent pain were again submitted to a feature selection, which resulted in the selection of seven items, of which six were from Spielberger's State–Trait Anxiety Inventory (STAI; blue columns in panel A). The line plot (panel B) shows the performance of different iterations of the reduced-set questionnaires for the prediction of persistent pain after breast cancer surgery. All possible sums of the selected items, which can take values of $n \in N$, $[0, \ldots, 3]$ for Beck's Depression Inventory (BDI) and Spielberger's State–Trait Anger Expression Inventory (STAXI-2), and $n \in N$, $[1, \ldots, 4]$ for STAI-state and STAI-trait, were iteratively tested with respect to their classification performance. The main criterion was the product of sensitivity and specificity. The green line shows the product of sensitivity and specificity for different item sums. Given are the medians of 1000 bootstrap resampling runs, surrounded by the 95% bootstrap confidence intervals (2.5th to 97.5th percentiles). The figure was created using the R software package (version 3.4.1 for Linux; http://CRAN.R-project.org/).[28]

sensations combined with insufficient strategies to calm those sensations in mind may develop a vicious circle that worsens both pain and mood.[39]

In agreement with what is already known, the short set of seven items seems to provide the key psychological parameters relevant for persistent pain after breast cancer treatments. The results suggest that psychological factors (i.e. characteristics of depressive mood and anxiety), assessed when the patients have already established persistent pain at 6 months, are sensitive in identifying patients with a long-term maladaptive outcome after breast cancer treatments. However, in line with other reported predictive factors (e.g. those reported by Sipilä and colleagues,[17] Lötsch and colleagues,[25] Meretoja and colleagues[46]), the association of persistent pain with psychological factors was only modest although both, the complete and the reduced lists of questions, had a high NPV. This indicates that the list of factors should be extended by other psychological characteristics (e.g. pain catastrophising,[47] inflexibility,[48] self-reported pain vigilance,[7,49,50] and resilience). In addition, biochemical, genetic, and other biomarkers[25] should be included.

In a previous attempt at reducing one of the included questionnaires, namely the STAI to a relevant item subset feasible for diagnostics, a correlative approach was taken combining the highly correlated items and selecting those that correlated best with the scores obtained using the full questionnaire. A reduced six-item set was proposed.[19] Applying this set on the present test data subset resulted in a sensitivity and specificity to detect patients with persistent pain of 47.6% and 67.7%, respectively, which corresponds to a balanced accuracy of correct patient group assignment of 57.6% and an NPV of persistent pain of 93.3%, which is slightly outperformed by the present results obtained with the two reduced subscales of STAI (Table 1) and the combined seven-item set (Table 3). The mentioned approach[19] focused on non-redundancy addressing highly correlated parameters. The present analysis used a machine-learning approach aiming at optimisation and performance of an algorithm.[30] Given the classification accuracy of the full questionnaires as a benchmark and considering the psychological plausibility of the proposed item subset, further feature selection methods on the present two-class problem (persistent or non-persistent pain), such as logistic regression or adaptive boosting,[51] were not tested as benchmarking of classifiers as families of item sets were not desired.

The psychological questionnaires were used based on the expectation that they provide information relevant for the identification of persistent post-surgical pain after breast

Table 2 Set of items that resulted from a feature selection analysis of the best suited items of each of the four questionnaires (BDI, STAXI-2, STAI-State, and STAI-Trait). The analysis identified seven items across the questionnaires that provide a similar prediction performance of persistent pain as the single complete or reduced scores. The resulting items originated mainly from the STAI; only one was from BDI and none from STAXI-2. After rescaling of the STAI-related responses into $n \in N, [0, \ldots, 3]$, a possible predictive tool would use the algorithm if $\sum_1^7 \text{items} < 7$ then a subject belongs to the 'non-persistent pain' group else to the 'persistent pain' group. BDI, Beck's Depression Inventory; STAI, State–Trait Anxiety Inventory; STAXI-2, Spielberger's State–Trait Anger Expression Inventory

Questionnaire	Item	Scaling	
		Original	Rescaled in proposed combined item set
BDI	#8: Self-criticalness	0, 1, 2, or 3	0, 1, 2, or 3
STAI-State	#1: I feel calm	1, 2, 3, or 4	0, 1, 2, or 3 (original rating − 1)
	#10: I feel comfortable		
	#20: I feel pleasant		
STAI-Trait	#21: I feel pleasant		
	#22: I get easily fatigued		
	#26: I feel rested		

Table 3 Test performance measures of a seven-item combined questionnaire. All items from each questionnaire that had been found most suitable for predicting persistent pain in previous analytical steps were again submitted to feature selection, which resulted in the final selection of seven items. These comprised one item from Beck's Depression Inventory (BDI), three items from State–Trait Anxiety Inventory (STAI)-State, and three items from STAI-Trait. The item reduction and creation of reduced questioners was performed on two-thirds of the data obtained via class-proportional random split of the original data set ('training data') whereas the assessment of predictive performance was performed with the one-third remaining data ('test data'). NPV, negative predictive value; PPV, positive predictive value

Test performance measure	Combined questionnaire
Sensitivity	0.61905
Specificity	0.65455
PPV	0.14607
NPV	0.94737
Balanced accuracy	0.63680

cancer surgery based on previous literature.[14–18] However, the results of the present analysis indicate that the main strength of these factors is the exclusion of patients at risk for developing persistent pain. This shows that machine learning and statistics are not identical. Machine learning focuses on the performance of algorithms, the clear focus being on the utility of the psychological information to identify the clinical course of pain. The statistical association between the questionnaire items and persistent pain, is reflected by the ability of the algorithm to associate the patients with the correct group of non-persistent or persistent pain that was better than guessing (approximately 60% balanced accuracy; Table 1), indicating that the questionnaire items contained relevant information about the time course of pain.

Conclusions

Using a data-driven machine-learning approach, a short list of seven items from BDI and Spielberger's STAI was selected as a basis for their association with the possible persistence of pain after breast cancer surgery. When compared with the complete forms of the questionnaires, the seven-item version offers a briefer and at least as accurate identification of women in whom pain persistence is unlikely (almost 95% NPV). Thus, the present results offer a shorter set of psychological data required to identify patients at risk for PPSP. Validity of the selected subset of items needs to be further tested in a different cohort of breast cancer patients and after other types of surgeries. These results suggest that machine learning is a promising approach for the development of predictive tools for persistent pain based on psychological factors.

Authors' contributions

Concept and design of experiments: E.K., R.S.
Concept and design of the analysis: J.L., R.S., E.K.
Data analysis: J.L.
Writing of the paper: J.L., R.S., E.K., V.D.
Data collection: E.K., R.S.
Critical revision of the manuscript for important intellectual content: E.K., R.S., V.D.

Declaration of interest

The authors have declared that no competing interests exist.

Funding

The work has been supported by the European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 602919 (GLORIA, E.K., J.L.) and no. 602891 (NeuroPain, E.K., R.S.). Additional support was received by the Landesoffensive zur Entwicklung wissenschaftlich-ökonomischer Exzellenz (LOEWE, J.L.), Zentrum: Translational Medicine and Pharmacology. The funders had no role in method design, data selection and analysis, decision to publish, or preparation of the manuscript.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.bja.2018.06.007.

References

1. Burke AL, Mathias JL, Denson LA. Psychological functioning of people living with chronic pain: a meta-analytic review. *Br J Clin Psychol* 2015; **54**: 345–60
2. Fillingim RB, Ohrbach R, Greenspan JD, et al. Psychological factors associated with development of TMD: the OPPERA prospective cohort study. *J Pain* 2013; **14**: T75–90
3. Linton SJ, Nicholas MK, MacDonald S, et al. The role of depression and catastrophizing in musculoskeletal pain. *Eur J Pain* 2011; **15**: 416–22

4. McCracken LM, Gross RT, Aikens J, Carnrike Jr CL. The assessment of anxiety and fear in persons with chronic pain: a comparison of instruments. *Behav Res Ther* 1996; **34**: 927–33
5. Knaster P, Estlander AM, Karlsson H, Kaprio J, Kalso E. Diagnosing depression in chronic pain patients: DSM-IV major depressive disorder vs. Beck Depression Inventory (BDI). *PLoS One* 2016; **11**. e0151982
6. Margari F, Lorusso M, Matera E, et al. Aggression, impulsivity, and suicide risk in benign chronic pain patients—a cross-sectional study. *Neuropsychiatr Dis Treat* 2014; **10**: 1613–20
7. Dimova V, Lötsch J, Hühne K, et al. Association of genetic and psychological factors with persistent pain after cosmetic thoracic surgery. *J Pain Res* 2015; **8**: 829–44
8. Lautenbacher S, Huber C, Schofer D, et al. Attentional and emotional mechanisms related to pain as predictors of chronic postoperative pain: a comparison with other psychological and physiological predictors. *Pain* 2010; **151**: 722–31
9. Bruce J, Thornton AJ, Scott NW, et al. Chronic preoperative pain and psychological robustness predict acute postoperative pain outcomes after surgery for breast cancer. *Br J Cancer* 2012; **107**: 937–46
10. Meretoja TJ, Leidenius MH, Tasmuth T, Sipila R, Kalso E. Pain at 12 months after surgery for breast cancer. *JAMA* 2014; **311**: 90–2
11. Vlaeyen JW, Linton SJ. Fear-avoidance model of chronic musculoskeletal pain: 12 years on. *Pain* 2012; **153**: 1144–7
12. Vlaeyen JW, Crombez G, Linton SJ. The fear-avoidance model of pain. *Pain* 2016; **157**: 1588–9
13. Edwards RR, Dworkin RH, Sullivan MD, Turk DC, Wasan AD. The role of psychosocial processes in the development and maintenance of chronic pain. *J Pain* 2016; **17**: T70–92
14. Pergolizzi JV, Gharibo C, Ho K-Y. Treatment considerations for cancer pain: a global perspective. *Pain Pract* 2015; **15**: 778–92
15. Dableh LJ, Yashpal K, Henry JL. Neuropathic pain as a process: reversal of chronification in an animal model. *J Pain Res* 2011; **4**: 315–23
16. Chou R, Gordon DB, de Leon-Casasola OA, et al. Management of postoperative pain: a clinical practice guideline from the American Pain Society, the American Society of Regional Anesthesia and Pain Medicine, and the American Society of Anesthesiologists' Committee on Regional Anesthesia, Executive Committee, and Administrative Council. *J Pain* 2016; **17**: 131–57
17. Sipilä R, Estlander A-M, Tasmuth T, Kataja M, Kalso E. Development of a screening instrument for risk factors of persistent pain after breast cancer surgery. *Br J Cancer* 2012; **107**: 1459–66
18. Schou Bredal I, Smeby NA, Ottesen S, Warncke T, Schlichting E. Chronic pain in breast cancer survivors: comparison of psychosocial, surgical, and medical characteristics between survivors with and without pain. *J Pain Symptom Manage* 2014; **48**: 852–62
19. Marteau TM, Bekker H. The development of a six-item short-form of the state scale of the Spielberger State–Trait Anxiety Inventory (STAI). *Br J Clin Psychol* 1992; **31**: 301–6
20. Spielberger CD. *Staxi-2: state–trait anger expression inventory-2; professional manual*. Odessa, FL: PAR, Psychological Assessment Resources; 1999
21. Kaunisto MA, Jokela R, Tallgren M, et al. Pain in 1,000 women treated for breast cancer: a prospective study of pain sensitivity and postoperative pain. *Anesthesiology* 2013; **119**: 1410–21
22. Trimble EL, Ungerleider RS, Abrams JA, et al. Neoadjuvant therapy in cancer treatment. *Cancer* 1993; **72**: 3515–24
23. Gagliese L, Weizblit N, Ellis W, Chan VW. The measurement of postoperative pain: a comparison of intensity scales in younger and older surgical patients. *Pain* 2005; **117**: 412–20
24. Macrae WA. Chronic pain after surgery. *Br J Anaesth* 2001; **87**: 88–98
25. Lötsch J, Ultsch A, Kalso E. Prediction of persistent post-surgery pain by preoperative cold pain sensitivity: biomarker development with machine-learning-derived analysis. *Br J Anaesth* 2017; **119**: 821–9
26. Beck AT, Ward CM, Mendelson M, Mock JE, Erbaugh JK. An inventory for measuring depression. *Arch Gen Psychiatr* 1961; **4**: 561–71
27. Spielberger CD, Gorsuch RL, Lushene RE. *The state–trait anxiety inventory: test manual*. Palo Alto, CA: Consulting Psychologists; 1970. p. 22
28. R Development Core Team. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing; 2008
29. Shalev-Shwartz S, Ben-David S. *Understanding machine learning: from theory to algorithms*. Cambridge University Press; 2014
30. Lotsch J, Ultsch A. Machine learning in pain research. *Pain* 2018; **159**: 623–30
31. Altman DG, Bland JM. Diagnostic tests: 1. Sensitivity and specificity. *BMJ* 1994; **308**: 1552
32. Altman DG, Bland JM. Diagnostic tests: 2. Predictive values. *BMJ* 1994; **309**: 102
33. Brodersen KH, Ong CS, Stephan KE, Buhmann JM. *The balanced accuracy and its posterior distribution. 20th International conference on pattern recognition (ICPR)* 2010. p. 3121–4
34. Efron B, Tibshirani RJ. *An introduction to the Bootstrap*. San Francisco: Chapman and Hall; 1995
35. Ho TK. *Random decision forests. Proceedings of the third international conference on document analysis and recognition*, vol. 1. IEEE Computer Society; 1995. p. 278
36. Breiman L. Random forests. *Mach Learn* 2001; **45**: 5–32
37. Ultsch A, Lötsch J. Computed ABC analysis for rational selection of most informative variables in multivariate data. *PLoS One* 2015; **10**. e0129767
38. Crombez G, Eccleston C, Van Damme S, Vlaeyen JW, Karoly P. Fear-avoidance model of chronic pain: the next generation. *Clin J Pain* 2012; **28**: 475–83
39. Vlaeyen JW, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. *Pain* 2000; **85**: 317–32
40. Keogh E, Asmundson GJG. Negative affectivity, catastrophizing and anxiety sensitivity. In: Asmundson GJG, Vlaeyen JW, Crombez G, editors. *Understanding and treating fear of pain*. Oxford: Oxford University Press; 2004. p. 91–115
41. Keogh E, Cochrane M. Anxiety sensitivity, cognitive biases, and the experience of pain. *J Pain* 2002; **3**: 320–9
42. Eccleston C, Crombez G. Pain demands attention: a cognitive–affective model of the interruptive function of pain. *Psychol Bull* 1999; **125**: 356–66
43. de Rooij A, de Boer MR, van der Leeden M, Roorda LD, Steultjens MPM, Dekker J. Cognitive mechanisms of change in multidisciplinary treatment of patients with

chronic widespread pain: a prospective cohort study. *J Rehabil Med* 2014; **46**: 173–80
44. Goubert L, Trompetter H. Towards a science and practice of resilience in the face of pain. *Eur J Pain* 2017; **21**: 1301–15
45. Khan RS, Ahmed K, Blakeway E, et al. Catastrophizing: a predictive factor for postoperative pain. *Am J Surg* 2011; **201**: 122–31
46. Meretoja TJ, Andersen KG, Bruce J, et al. Clinical prediction model and tool for assessing risk of persistent pain after breast cancer surgery. *J Clin Oncol* 2017; **35**: 1660–7
47. Sullivan MJL, Bishop SR, Pivik J. The pain catastrophizing scale: development and validation. *Psychol Assess* 1995; **7**: 524–32
48. Wicksell RK, Renofalt J, Olsson GL, Bond FW, Melin L. Avoidance and cognitive fusion—central components in pain related disability? Development and preliminary validation of the Psychological Inflexibility in Pain Scale (PIPS). *Eur J Pain* 2008; **12**: 491–500
49. Quartana PJ, Campbell CM, Edwards RR. Pain catastrophizing: a critical review. *Expert Rev Neurother* 2009; **9**: 745–58
50. Horn-Hofmann C, Scheel J, Dimova V, et al. Prediction of persistent post-operative pain: pain-specific psychological variables compared with acute post-operative pain and general psychological variables. *Eur J Pain* 2018; **22**: 191–202
51. Schapire RE, Freund Y. A short introduction to boosting. *J Jap Soc Artif Intell* 1999; **14**: 771–80

Handling editor: L. Colvin