



Support Vector Machine Analysis of Functional Magnetic Resonance Imaging of Interoception Does Not Reliably Predict Individual Outcomes of Cognitive Behavioral Therapy in Panic Disorder with Agoraphobia

Benedikt Sundermann^{1*†}, Jens Bode^{1†}, Ulrike Lueken^{2,3}, Dorte Westphal², Alexander L. Gerlach⁴, Benjamin Straube⁵, Hans-Ulrich Wittchen², Andreas Ströhle⁶, André Wittmann⁶, Carsten Konrad^{5,7}, Tilo Kircher⁵, Volker Arolt⁸ and Bettina Pfleiderer^{1,9}

OPEN ACCESS

Edited by:

Joaquim Radua,
Fidmag Sisters
Hospitalers, Spain

Reviewed by:

Gemma C. Monté Rubio,
Fundació ACE, Spain
Hugo Schnack,
Utrecht University, Netherlands

*Correspondence:

Benedikt Sundermann
benedikt.sundermann@
uni-muenster.de

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Neuroimaging and Stimulation,
a section of the journal
Frontiers in Psychiatry

Received: 05 April 2017

Accepted: 22 May 2017

Published: 09 June 2017

Citation:

Sundermann B, Bode J, Lueken U,
Westphal D, Gerlach AL, Straube B,
Wittchen H-U, Ströhle A, Wittmann A,
Konrad C, Kircher T, Arolt V and
Pfleiderer B (2017) Support Vector
Machine Analysis of Functional
Magnetic Resonance Imaging of
Interoception Does Not Reliably
Predict Individual Outcomes of
Cognitive Behavioral Therapy in
Panic Disorder with Agoraphobia.
Front. Psychiatry 8:99.
doi: 10.3389/fpsy.2017.00099

¹Department of Clinical Radiology, University Hospital Münster, Münster, Germany, ²Department of Psychology, Institute of Clinical Psychology and Psychotherapy, Technische Universität Dresden, Dresden, Germany, ³Center for Mental Health, Department of Psychiatry, Psychosomatics, and Psychotherapy, University Hospital Würzburg, Würzburg, Germany, ⁴Klinische Psychologie und Psychotherapie, Universität zu Köln, Cologne, Germany, ⁵Department of Psychiatry and Psychotherapy, Philipps University of Marburg, Marburg, Germany, ⁶Department of Psychiatry and Psychotherapy, Charité – University Medicine Berlin, Berlin, Germany, ⁷Department of Psychiatry and Psychotherapy, Agaplesion Diakonieklinikum Rotenburg, Rotenburg, Germany, ⁸Department of Psychiatry and Psychotherapy, University Hospital Münster, Münster, Germany, ⁹Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Münster, Münster, Germany

Background: The approach to apply multivariate pattern analyses based on neuro imaging data for outcome prediction holds out the prospect to improve therapeutic decisions in mental disorders. Patients suffering from panic disorder with agoraphobia (PD/AG) often exhibit an increased perception of bodily sensations. The purpose of this investigation was to assess whether multivariate classification applied to a functional magnetic resonance imaging (fMRI) interoception paradigm can predict individual responses to cognitive behavioral therapy (CBT) in PD/AG.

Methods: This analysis is based on pretreatment fMRI data during an interoceptive challenge from a multicenter trial of the German PANIC-NET. Patients with DSM-IV PD/AG were dichotomized as responders ($n = 30$) or non-responders ($n = 29$) based on the primary outcome (Hamilton Anxiety Scale Reduction $\geq 50\%$) after 6 weeks of CBT (2 h/week). fMRI parametric maps were used as features for response classification with linear support vector machines (SVM) with or without automated feature selection. Predictive accuracies were assessed using cross validation and permutation testing. The influence of methodological parameters and the predictive ability for specific interoception-related symptom reduction were further evaluated.

Results: SVM did not reach sufficient overall predictive accuracies (38.0–54.2%) for anxiety reduction in the primary outcome. In the exploratory analyses, better accuracies (66.7%) were achieved for predicting interoception-specific symptom relief as an alternative outcome domain. Subtle information regarding this alternative response criterion but not the primary outcome was revealed by *post hoc* univariate comparisons.

Conclusion: In contrast to reports on other neurofunctional probes, SVM based on an interoception paradigm was not able to reliably predict individual response to CBT. Results speak against the clinical applicability of this technique.

Keywords: panic disorder, agoraphobia, cognitive behavioral therapy, interoception, functional magnetic resonance imaging, diagnostic classification, machine learning, support vector machines

INTRODUCTION

Applying multivariate pattern analysis (MVPA) techniques from the field of machine learning to functional magnetic resonance imaging (fMRI) data has been proposed as a strategy to develop diagnostic or predictive tools for mental disorders. MVPA integrates potentially discriminative information from multiple brain locations, states, or imaging modalities instead of analyzing them in isolation. MVPA techniques are applied to learn decision rules (classifiers) based on labeled training data. These rules are subsequently applied to diagnostically label previously unseen data (1–8). Particularly, support vector machines (SVM) have been used to extract meaningful information from noisy and high-dimensional fMRI data (5, 6). Predicting individual therapeutic outcomes in psychiatric patients (including those with anxiety disorders) is an emergent focus of these methodological efforts (9–15). Responses to pharmacotherapy and to psychotherapy have been proposed to be mediated by neurobiological factors (12, 14, 16–21). Therefore, predictive neuroimaging biomarkers are of particular scientific interest as candidate tools to guide clinical treatment decision in individual psychiatric patients (9, 14, 22).

Cognitive behavioral therapy (CBT) alone or in combination with pharmacotherapy is the first-line treatment of choice in patients suffering from panic disorder with agoraphobia (PD/AG) (23–29). While there is a clinically significant proportion of non-responders (26, 28, 30), there is only incipient but rapidly increasing knowledge of moderators and predictors of CBT response (12, 31, 32). Neural correlates of CBT in PD/AG have been investigated in fMRI group comparison studies suggesting a role of altered neural activity in networks regulating negative emotions as well as fear conditioning and extinction (12, 18, 21, 33–36). Recently, first reports of successful applications of MVPA to predict individual CBT outcomes in PD/AG based on fMRI emerged: Hahn et al. reported an overall diagnostic accuracy of up to 82% using Gaussian process classifiers (GPCs) in a meta-learning scheme to train models based on task fMRI data from a fear condition paradigm in an overlapping sample (11). Ball et al. utilized random forest classification based on an emotion regulation task. They reached accuracies of 79% in a mixed sample of PD and generalized anxiety disorder and 85% in the PD subsample (37).

An intensified and abnormal internal focus of attention to bodily sensations (interoception) is a characteristic feature observed in PD/AG (38, 39). This comprises increased self-report of bodily symptoms (particularly cardiac) as well as their dysfunctional cognitive appraisal including catastrophizing (39). Interoception is therefore assumed to be an important determinant of maintenance of PD/AG (28, 38, 39). It is thus specifically

addressed by CBT *via* interoceptive exposure (28). Interoception can be effectively studied by fMRI and is associated with activity in a widespread cerebral network overlapping with established fear circuitry models (39–42).

The purpose of this investigation was to assess whether fMRI based on an interoception task acquired at multiple sites combined with SVM, a well-established MVPA technique, can predict CBT response of individual patients with PD/AG. Beyond that, we aimed at further exploring the influence of methodological decisions and the predictive ability for specific interoception-related symptom reduction. We additionally performed univariate group analyses comparing responders and non-responders to assess feature set information content and to evaluate the general suitability of the paradigm to detect neural processes related to therapy response.

MATERIALS AND METHODS

Subjects

This investigation is based on fMRI data in a subset of patients of the multicenter, randomized-controlled trial “Mechanism of action in CBT” (MAC) (43) within the framework of the German research network PANIC-NET (44). Primary goal of the MAC trial is to identify mechanisms through which CBT achieves its beneficial effects as well as mediators and moderators of response. It involves the acquisition of a broad spectrum of clinical, behavioral, physiological, experimental, and genetic data. Written informed consent was obtained from all participants in accordance with the Declaration of Helsinki. The randomized clinical trial (isrctn.org identifier: ISRCTN80046034) was approved by the ethics committee of the Medical Faculty of the Technische Universität Dresden (agreement EK 164082006). The neuroimaging components were approved by the ethics committee of the Medical Faculty of the RWTH Aachen University, Aachen (agreement EK 073/07) and at all local sites (43).

The overall MAC sample involved adult outpatients ($n = 369$) who met criteria for a current primary diagnosis of PD/AG (43) according to DSM-IV-TR (45). Only moderate exclusion criteria were adopted to allow for typical comorbidity seen in routine care. They comprised comorbid psychotic or bipolar I disorders, current substance dependence or abuse, a current suicidal intent, borderline personality disorder, ongoing psychotherapeutic or psychopharmacological treatment as well as procedure-specific contraindications. Psychometric assessments in participants of the fMRI substudy included the Hamilton Scale for Anxiety (HAM-A) (46, 47), Beck Depression Inventory (BDI-II) (48), Anxiety Sensitivity Index (49), Clinical Global Impression (50), PD/AG Scale (51), trail-making task (52), and digit span task

from the German Wechsler Adult Intelligence Scale—Revision IV (53). Patients received manualized exposure-based CBT encompassing 12 × 100 min treatment sessions (two subgroups either with or without therapist-guided exposure) or were allocated to a wait-list control group (data not used in this analysis). Please see the MAC methods paper for full details on patient recruitment, treatment, and data collection (43). A subgroup of patients ($n = 89$) were invited to participate in the fMRI substudy (33). This analysis is based on a subsample ($n = 59$) representing all PD/AG patients who completed CBT (including the assessment of clinical outcomes) as well as an interoception fMRI task (41) at baseline and fMRI data quality assessment (see Figure S1 in Supplementary Material for a flowchart of patient selection). Analogous to Hahn et al. (11) a reduction in HAM-A scores $\geq 50\%$ (primary outcome) from baseline to posttreatment assessment was used as a standard criterion for treatment response (43). Data from responders ($n = 30$) and non-responders ($n = 29$) were analyzed here. Demographic and clinical details of responders and non-responders are presented in **Table 1**. Statistical assessment of clinical and demographical data was accomplished using IBM SPSS Statistics (version 22, IBM, Armonk, NY, USA, RRID:SCR_002865).

A secondary response criterion was available in a subgroup of 54 out of 59 patients (see Multivariate Classification on exploratory analyses): this was based on the assessment of the intensity of bodily symptoms (SI) and experienced anxiety (EA) during two sessions of “interoceptive exposure” to bodily sensations during CBT (43, 54). Self-report data on SI and EA were documented on a scale from 0 to 10 before (fourth CBT session) and after (fifth CBT session) an interoceptive exercise (IE) involving repeated self-guided exposure at home. The interoceptive exposure involved a wide range of bodily stimuli. To calculate the summary “interoceptive” score used here, the summary measures of the three stimuli with largest effect sizes: breathing through a straw, rotating around the longitudinal body axis, and hyperventilation were chosen (54). Absolute between-session differences of SI and EA scores were averaged over these three stimuli to obtain cumulated measures of SI or EA reduction in individual patients. Patients with a cumulated SI and EA reduction above or below the group mean were classified as responders and non-responders, respectively. This alternative response criterion resulted in a further subgrouping of IE responders ($n = 26$) and IE non-responders ($n = 28$). Interoception-based response was not associated with overall response based on 50% HAM-A reduction ($\chi^2 = 0.297$, $p = 0.586$). See Table S1 in Supplementary Material for clinical and demographical characteristics of IE responders and non-responders.

fMRI Data Acquisition and First-Level Analyses

Data acquisition was accomplished at four imaging centers using T2*-weighted gradient echo echo-planar imaging (225 volumes, 36 slices, matrix 64×64 , field of view 210 mm, reconstructed as $3.6 \text{ mm} \times 3.6 \text{ mm} \times 3.6 \text{ mm}$ voxels, echo time 35 ms, repetition time 3,000 ms, and flip angle 90°) at 3 T (Aachen and Münster: Achieva, Philips, Best, Netherlands; Berlin: GE Healthcare, Little

Chalfont, UK; Dresden: Magnetom Trio, Siemens, Erlangen, Denmark).

During the fMRI scan, participants performed a mental tracking paradigm (41) adapted from the so-called “Schandry task” (55). The paradigm examines the effects of focusing one’s attention internally (interoception) vs. externally (exteroception) using a block design with four blocks. External stimuli were identical in both conditions: hard to hear clicking sounds ($n = 104$ per block): during exteroception, participants were instructed to silently count the clicking sounds and to subsequently report the number of clicks. During interoception, participants were instructed to silently count their own heartbeats and to report the number of heartbeats counted in a particular interval. Subjects were randomized to either start with an interoception or exteroception block (see **Table 1**). In total, two blocks of interoception (I1, I2) and two blocks of exteroception (E1, E2) were presented. This paradigm had been validated previously in anxiety sensitive females (41).

Preprocessing and first-level analyses were conducted with SPM5¹ (RRID:SCR_007037). Images were realigned, normalized and resliced (voxel size $2 \text{ mm} \times 2 \text{ mm} \times 2 \text{ mm}$), and smoothed with a Gaussian kernel (full width at half maximum: 8 mm). Movement correction parameters were used as regressors in the first-level model. Data were filtered with a high-pass filter (cutoff period of 128 s). The two blocks of interoception and the two blocks of exteroception, respectively, were added to build one condition ($I = I1 + I2$; $E = E1 + E2$) (41).

Multivariate Classification

Maps representing either the simple contrast “interoception” (I) (i.e., the respective beta-map) or the differential contrast “interoception > exteroception” ($I > E$) in individual subjects were used as features for subsequent multivariate classification and *post hoc* univariate group comparisons.

General Approach and Hypothesis Tests

Modeling and validation were implemented using the Machine learning Application for NeuroImaging Analyses (MANIA, version 2.5) (56). Diagnostic performance was assessed using leave-one-out cross-validation (57). The statistical significance of inferentially tested methods was estimated using permutation testing (58) with 100 permutations.

Support vector machine models tested here were based on soft-margin support vector classification (C-SVC) from LIBSVM (59). In SVM models, a hyperplane is defined in order to distinguish between responders and non-responders. Models are optimized using a kernel by maximizing the margin of separation between groups based on the datasets closest to the hyperplane. Model parameters can be chosen to adjust model complexity. In this particular case, the penalty-term C adjusts the models’ tolerance for misclassifications in the training dataset (57, 60–62).

Support vector machines can be combined with different methods for dimensionality reduction and feature selection (FS) with the aim to improve diagnostic accuracies (57, 63). In this

¹<http://www.fil.ion.ucl.ac.uk/spm>.

TABLE 1 | Basic characteristics of responders and non-responders to cognitive behavioral therapy in the main analysis (primary outcome, responder-threshold: 50% HAM-A reduction compared to pretreatment baseline).

| | | Responders | Non-responders | Test statistic | df | p |
|--------------------------------------|------------------|-------------|----------------|------------------|------|-----------------------|
| Number | | 30 | 29 | | | |
| Sex | Female | 18 (60.0%) | 19 (65.5%) | $\chi^2 = 0.192$ | 1 | 0.661 ^c |
| | Male | 12 (40.0%) | 10 (34.5%) | | | |
| Age (years) | | 36.8 ± 12.2 | 37.3 ± 10.1 | $t = -0.198$ | 57 | 0.844 ^a |
| Education | Lower secondary | 13 (43.3%) | 18 (62.1%) | $\chi^2 = 2.990$ | 2 | 0.212 ^c |
| | Higher secondary | 11 (36.7%) | 9 (31.0%) | | | |
| | University | 6 (20.0%) | 2 (6.9%) | | | |
| Site (n) | Aachen | 0 (0.0%) | 2 (6.9%) | $\chi^2 = 4.140$ | 3 | 0.247 ^c |
| | Berlin | 11 (36.7%) | 7 (24.1%) | | | |
| | Dresden | 10 (33.3%) | 14 (48.3%) | | | |
| | Münster | 9 (30.0%) | 6 (20.7%) | | | |
| CBT arm | Therapist-guided | 14 (46.7%) | 19 (65.5%) | $\chi^2 = 2.126$ | 1 | 0.192 ^c |
| | Non-guided | 16 (53.3%) | 10 (34.5%) | | | |
| Randomized first fMRI condition (n) | Interoception | 17 (56.7%) | 11 (37.9%) | $\chi^2 = 2.076$ | 1 | 0.150 ^c |
| | Exteroception | 13 (43.3%) | 18 (62.1%) | | | |
| HAM-A | Before CBT | 24.0 ± 5.5 | 25.1 ± 5.3 | $t = -0.784$ | 57 | 0.436 ^a |
| | After CBT | 7.9 ± 3.3 | 18.1 ± 5.2 | $t = -8.996$ | 46.7 | <0.001 ^{a,*} |
| BDI-II | Before CBT | 15.9 ± 9.9 | 17.0 ± 7.9 | $t = -0.469$ | 57 | 0.641 ^a |
| | After CBT | 6.5 ± 5.3 | 12.9 ± 8.8 | $t = -3.361$ | 45.4 | 0.002 ^{a,*} |
| ASI | Before CBT | 30.9 ± 9.7 | 31.0 ± 12.1 | $t = -0.025$ | 57 | 0.980 ^a |
| | After CBT | 12.9 ± 6.8 | 18.8 ± 10.4 | $t = -2.580$ | 57 | 0.012 ^{a,*} |
| CGI | Panic symptoms | 5 (4–7) | 5 (4–7) | | | 0.570 ^b |
| | Anxiety | 3 (1–6) | 4 (3–5) | | | 0.009 ^{b,*} |
| PAS | | 21.0 ± 8.2 | 29.6 ± 6.2 | $t = -4.553$ | 57 | <0.001 ^{a,*} |
| TMT (s) | A | 26.3 ± 9.5 | 27.1 ± 8.3 | $t = -0.337$ | 57 | 0.737 ^a |
| | B | 59.6 ± 20.1 | 58.5 ± 17.7 | $t = 0.210$ | 57 | 0.834 ^a |
| Digit span task | Total | 15.1 ± 2.8 | 14.2 ± 3.1 | $t = 1.152$ | 57 | 0.254 ^a |
| Comorbid depression (n) ^d | Before CBT | 10 (33.3%) | 9 (31.0%) | $\chi^2 = 0.036$ | 1 | 0.850 ^c |
| | After CBT | 1 (3.7%) | 6 (20.7%) | $\chi^2 = 4.248$ | 1 | 0.039 ^{c,*} |

All test results without further specification were obtained at the first visit at base line assessment of CBT and represent "number (percentage)," "mean ± SD" or "median (range)."

^adenotes statistical significance ($p < 0.05$).

^a t -test.

^bMann–Whitney U -test.

^c χ^2 -test.

^dbased on BDI-II scores.

CBT, cognitive behavioral therapy; HAM-A, Hamilton Scale for Anxiety; BDI-II, Beck Depression Inventory II; ASI, Anxiety Sensitivity Index; CGI, Clinical Global Impression; PAS, PD/AG Scale; TMT, trail-making task.

study, whole-brain datasets were masked to reduce dimensionality and preselect features (64). Additionally, models were tested either without any automated FS, with a simple filter (based on results of a two-sample t -test) or with recursive feature elimination using linear SVMs (SVM-RFE). SVM-RFE is an iterative procedure in which unimportant features are removed based on their SVM weights. Compared to the simple filters, SVM-RFE takes dependencies among features into account (65). Here, we used the greedy approach to SVM-RFE (56, 66). The effects of FS on an exemplary feature set are illustrated in **Figure 1**.

We performed hypothesis-based tests on six different models with standard methodological choices to test whether these models were able to identify individual responders and non-responders (50% HAM-A reduction) based on pretreatment fMRI: Contrast-maps ($I, I > E$) were downsampled to a voxel size

of 4 mm × 4 mm × 4 mm using SPM representing the rounded original voxel size in order to limit feature set dimensionality by avoiding redundancies only introduced during image registration. The Automated Anatomical Labeling (AAL) atlas (67) resampled to an identical resolution was used as a mask to restrict analyses predominantly to voxels representing gray matter. For each of the two contrasts ($I, I > E$), we assessed modeling without further automated FS, with a t -test filter and with SVM-RFE. For models with FS, the number of features to be selected was $n = 4,557$ (20% of voxels within the AAL mask). We did not further restrict the number of features here to avoid models based on only one or few coherent brain region comprising multiple voxels. For all six linear C-SVC models, we chose an intermediate value of the penalty term ($C = 1$), the default setting in LIBSVM.

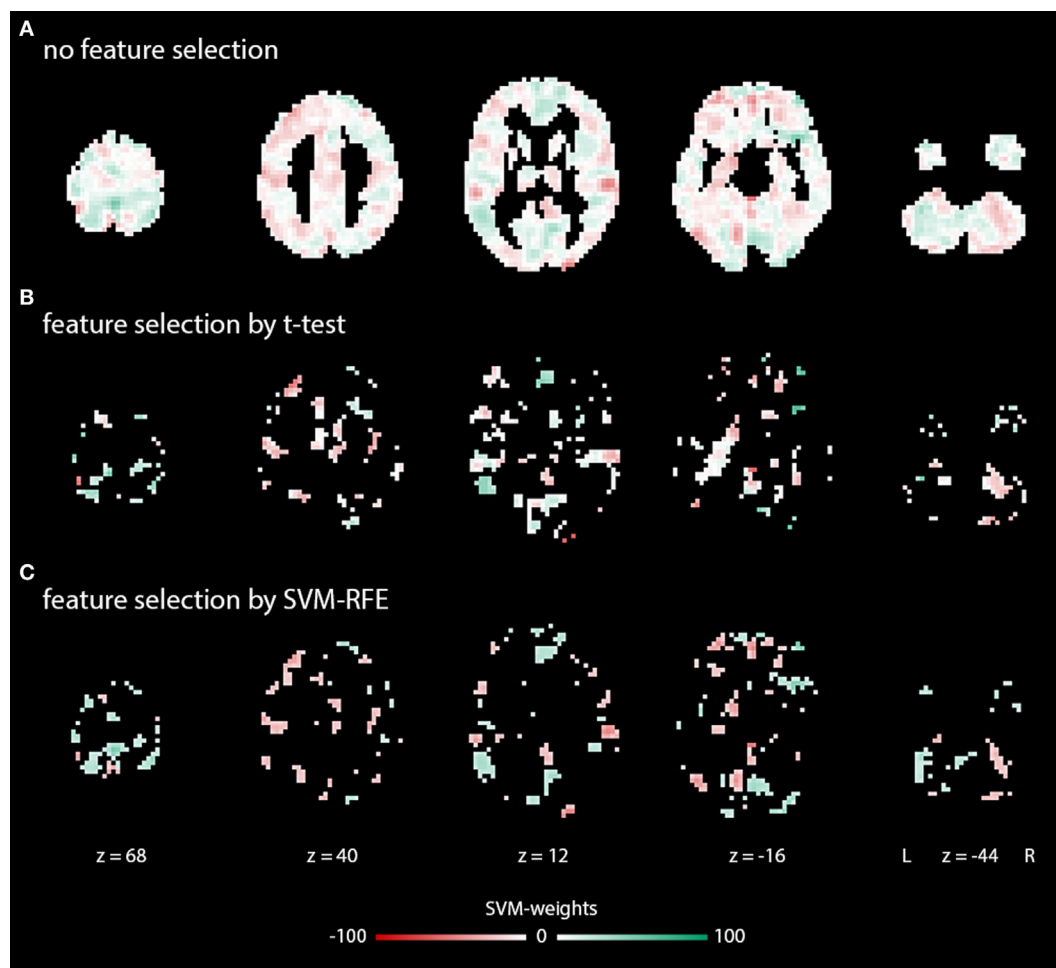


FIGURE 1 | Exemplary weight maps illustrating typical feature sets in the main analysis: **(A)** without feature selection (FS), **(B)** with FS by a *t*-test-based filter, **(C)** with FS by SVM-RFE. Illustrations were created using Mango (<http://ric.uthscsa.edu/mango/>).

Further Exploratory Analyses

Further potentially influencing factors were tested separately in exploratory analyses to identify factors that may improve diagnostic performance: (i) to assess whether interoception-related symptom relief would be a better predictable therapy response by SVM based on this fMRI interoception paradigm, we used the results of the alternative response criterion taking into account the response to an interoceptive challenge (IE responders vs. IE non-responders) (54). This contrasts with the standard HAM-A-based response criterion which focusses on clinically relevant general anxiety reduction. (ii) Different methodological decisions (feature set preparation, classification algorithm, modeling parameters, and software): the cost parameter was varied systematically $C = \{0.01, 1, 100\}$ for all analogous analyses with the main hypothesis tests and for all tests described hereafter. Another automated FS method was evaluated: selection by SVM weights. The influence of further gray matter masks was assessed: the cortical or subcortical Harvard-Oxford atlas (68–71) and a combined mask including the bilateral amygdala and parahippocampal gyrus as defined by Talairach labels (72, 73) in the

WFU PickAtlas toolbox.² The voxel size was kept either at its level after preprocessing ($2 \text{ mm} \times 2 \text{ mm} \times 2 \text{ mm}$) or further increased to $6 \text{ mm} \times 6 \text{ mm} \times 6 \text{ mm}$ in order to modify the feature set dimensionality. When different voxel resolutions were combined with FS, the number of features to be selected was adjusted to constantly reflect 20% of voxel within in the respective resampled mask. The Pattern Recognition for Neuroimaging Toolbox (PRoNTo, version 1.1, RRID:SCR_006908) (64) was used instead of MANIA. Both toolboxes rely on C-SVC from LIBSVM. However, there are some conceptual differences: while MANIA adopts the original linear C-SVC implementation from LIBSVM, PRoNTo uses precomputed linear kernels that are subsequently passed to LIBSVM. Both toolboxes are relatively new scientific software. We thus also wanted to exclude that simple data handling issues significantly influence MVPA results. Finally, we tested GPCs based on the Gaussian Processes for Machine Learning (74) implementation in MANIA (mean function: mean

²<http://fmri.wfubmc.edu/software/pickatlas>.

zero, covariance function: linear, likelihood function Likert). All analysis steps and parameters not explicitly mentioned here were identical with the inferentially tested models.

Post Hoc Univariate Group Comparisons

In order to assess the information content of the fMRI data to interpret multivariate classification performance, we additionally conducted a conventional univariate whole-brain analysis of response effects after therapy in patients. These were carried out with SPM. The “full factorial” design option was used with the factor “response to therapy” (yes, no) as independent factor. Separate analyses of variance were performed for both, the simple contrast I and the differential contrast I > E for both response criteria (HAM-A reduction and interoception criterion). The main effect of response was assessed. Results were cluster-size corrected for multiple comparisons on the cluster level at $p < 0.05$. This was equivalent to $p < 0.001$ with a minimum cluster extent of $k = 42$ contiguous resampled voxels with our given scanning parameters as estimated based on a Monte Carlo simulation implemented in Matlab (75).

RESULTS

Performance of Inferentially Tested Standard Approaches

Tested models did not yield significant diagnostic accuracies to identify individual responders and non-responders (50% HAM-A reduction) based on pretreatment fMRI with an interoception task. Overall accuracies ranged from 39.0 to 54.2% with sensitivities from 30.0 to 50.0% and specificities from 37.9 to 58.6%. See Table 2 for detailed results.

Exploratory Analyses

Further exploratory analyses aiming at methodological factors (feature set preparation, classification algorithm, modeling parameters, and software) did also not yield above-chance diagnostic performance. Overall accuracies ranged from 33.9 to 54.2% with sensitivities from 30.0 to 60.0% and specificities from 27.6 to 58.6%. An overview of models tested and detailed results are presented in Tables S2 and S3 in Supplementary Material.

Analyses with an alternative response criterion specifically aiming at a reduction of symptoms directly related to interoception (IE responders vs. IE non-responders) exhibited higher diagnostic accuracies compared with the standard response criterion. Overall accuracies ranged from 50.0 to 66.7% with sensitivities from 50.0 to 69.2% and specificities from 46.4 to 67.9%. See Table 3 for detailed results.

Post Hoc Univariate Group Comparisons

Conventional univariate group comparisons revealed a main effect of therapy response only for the interoception-specific response criterion with the contrast I > E (Figure S2 in Supplementary Material). No significant effects were observed for the interoception-specific outcome with the simple contrast I or for the HAM-A based primary outcome (contrasts I and I > E).

DISCUSSION

Diagnostic modeling based on a pretreatment interoception task with standard fMRI and voxel-wise SVM including FS did not achieve significant accuracies to predict individual CBT response in a randomized, controlled multicenter study. Using these methodological choices, we could not reach diagnostic performances

TABLE 2 | Results of main classification approaches to predict general anxiety reduction after CBT (hypothesis tests).

| GLM contrast | Feature selection (FS) | Accuracy (p) | Sensitivity (%) | Specificity (%) |
|--------------|------------------------|------------------|-----------------|-----------------|
| I | – | 39.0% (0.89) | 36.7 | 41.4 |
| I | <i>t</i> -Test filter | 39.0% (0.91) | 40.0 | 37.9 |
| I | SVM-RFE | 39.0% (0.91) | 36.7 | 41.4 |
| I > E | – | 39.0% (0.89) | 30.0 | 48.3 |
| I > E | <i>t</i> -Test filter | 54.2% (0.33) | 50.0 | 58.6 |
| I > E | SVM-RFE | 42.4% (0.79) | 40.0 | 44.8 |

Models based on C-SVC ($C = 1$). FS to select 20% of voxels. Statistical significance assessed by permutation testing. INT, interoception; EXT, exteroception; SVM-RFE, recursive feature elimination using support vector machines.

TABLE 3 | Results of an exploratory analysis with an interoception-specific response criterion (prediction of a reduction of bodily symptoms and anxiety during an interoceptive task, IE-responders vs. IE-non-responders).

| GLM contrast | Feature selection (FS) | Accuracy (p) | Sensitivity (%) | Specificity (%) |
|--------------|------------------------|------------------|-----------------|-----------------|
| I | – | 50.0% (0.48) | 50.0 | 50.0 |
| I | <i>t</i> -Test filter | 63.0% (0.14) | 69.2 | 57.1 |
| I | SVM-RFE | 48.2% (0.55) | 50.0 | 46.4 |
| I > E | – | 57.4% (0.21) | 57.7 | 57.1 |
| I > E | <i>t</i> -Test filter | 66.7% (0.02) | 65.4 | 67.9 |
| I > E | SVM-RFE | 57.4% (0.15) | 57.7 | 57.1 |

p-Values are reported only in order to exemplify the relationship between the observed accuracies and the distribution of chance level accuracies, but do not reflect planned hypothesis tests.

Models based on C-SVC ($C = 1$). FS to select 20% of voxels. Statistical significance assessed by permutation testing. INT, interoception; EXT, exteroception; SVM-RFE, recursive feature elimination using support vector machines.

of alternative models with different fMRI tasks. Such models have been reported by Hahn et al. in a fear conditioning fMRI paradigm (Gaussian process classifier in a meta-learning scheme) in an overlapping patient cohort within the MAC trial (11) and by Ball et al. in an emotion regulation task (37).

To interpret this negative finding regarding diagnostic accuracy for the primary endpoint (HAM-A reduction), it would be desirable to assess the following two questions separately: (1) Does the task fMRI data set convey sufficient information about the diagnostic question of interest? (2) Are the feature extraction and classification methods suitable to derive sufficiently powerful diagnostic models based on that information? In SVM as well as in MVPA in general, these two aspects are highly interconnected (patterns in MVPA are truly multivariate representations and conceptually different from univariate results in standard fMRI group analyses) (64, 76). Therefore, the following observation needs to be interpreted with caution: In *post hoc* univariate whole-brain analyses, we observed a significant main effect of response only for the alternative interoception-based response criterion with the differential contrast ($I > E$) (Figure S2 in Supplementary Material). This was the feature definition with which the best performance for diagnostic classification was achieved as well (see Exploratory Analyses). Therefore, in this study, the accuracy of SVM models followed the effects seen in univariate analyses. This may indicate that the interoception task did generally not yield sufficient information about the HAM-A based general response (primary outcome). However, one has to keep in mind that multivariate classification models can, in principle, utilize subthreshold information from multiple voxels, but even highly significant univariate group-level results do not guarantee high classification accuracies (2).

Thus, these negative findings may indicate that fear conditioning and extinction (11) as well as emotion regulation (37) may better reflect neural mechanisms involved in CBT and may therefore be more suitable to serve as a predictive tool than an interoceptive task, especially if measures of general anxiety are used as a reference standard for response. The fear conditioning paradigm has also been used to differentiate between PD/AG with and without depressive comorbidity (77).

It is not possible to identify unequivocally the reasons why our diagnostic approach has failed, nonetheless, we believe that it is important to report this negative finding since the methodologically diverse field of diagnostic MVPA in mental disorders is particularly susceptible to publication bias (7, 78, 79). Most importantly, interoceptive accuracy is still considered one of the major factors contributing to the development and maintenance of panic disorder (80).

Compared to other mental disorders (2, 7), few diagnostic MVPA studies have been reported to predict therapy outcomes in PD/AG so far (11, 37). Consequently, only a minority of available methods has been probed in this scenario. No consensus has yet been reached in the field as to which modeling techniques should be preferred (2, 6, 7). This is the first application of whole-brain voxel-based classification in this setting. Voxel-based approaches have been widely used in successful diagnostic modeling in other mental disorders (2, 7) and are commonly used for MVPA of fMRI data beyond diagnostic classification, particularly in combination

with SVM (64). SVM are particularly suitable for classification in datasets with a high dimensionality (i.e., number of features) compared to the number of observations (i.e., patients) (81–83). We have applied a combination of techniques for dimensionality reduction (63) with the aim to improve diagnostic performance: (1) imaging data were downsampled to a lower resolution reducing the number of feature by a factor of 1/8 compared to the original preprocessed data, without expected relevant information loss given the original acquisition resolution and data smoothness. (2) Images were masked to exclude voxels outside gray matter. (3) We applied automated FS to only include the 20% most relevant voxels. Literature-based definition of regions of interest (81) was, however, not feasible since no sufficiently reliable prior knowledge for this purpose was available in the literature. We thus selected features based on the dataset itself. Please note that FS was strictly included in the CV to avoid circular reasoning (84). Failure to do that, for example, selection of regions of interest after interpreting univariate group analyses in the same full dataset, is unfortunately a commonly observed mistake leading to overly optimistic estimates of diagnostic accuracies (7). *Post hoc* univariate group effects support the voxel-based approach with automatic FS: effects were only observed in few regions (see *Post Hoc* Univariate Group Comparisons; Figure S2 in Supplementary Material), so that the inclusion of finer-scale within-region patterns, which is possible in voxel-based modeling, seems preferable compared to large-scale inter-regional MVPA.

In addition to inferentially testing these established whole-brain SVM approaches, we explored the influence of methodological choices regarding feature set preparation, modeling parameters, and software. Furthermore, we assessed GPCs as an alternative to SVMs. However, we only observed minor differences in overall diagnostic accuracies and even none of these models reached clinically meaningful diagnostic performance. We thus conclude that the insufficient diagnostic accuracies observed in the hypothesis tests is not caused by insufficient methodological decisions within this range, highlighting the importance of fMRI task choice.

We also explored the diagnostic ability of our multivariate classification approach to predict reduction of symptoms directly related to interoceptive exposure. Better diagnostic accuracies were achieved with this alternative response criterion, particularly with the differential contrast ($I > E$) combined with simple automated FS (Table 3). Contrasts are used to test specific effects in general linear models of fMRI data (85). The differential contrast is aimed to be more specific for interoception than the simple contrast I as it excludes baseline effects and effects common to both task conditions. Though not directly amenable to statistical testing, results indicate that defining such specific differential contrasts facilitates improved feature extraction compared to simple contrasts (i.e., beta-maps). However, this is an exploratory result without independent validation and still no clinically useful diagnostic accuracies were achieved.

Some limitations apply to this work: subjects were stratified as responders or non-responders. Theoretically, response to therapy can be treated as a regression problem (86) instead of classification, particularly as some non-responders also experienced clinically relevant anxiety reduction. Regression requires

larger samples than classification to achieve sufficient statistical power. This is why we favored a classification in this pilot analysis.

The alternative response criterion based on interoception-related symptom reduction relies on self-report data and an IE at home. Thus, there is no external control of symptom relief and patient's adherence to the task. We can thus not exclude a social desirability bias in the self-report data (87). Results based on this alternative response criterion as well as results of additional methodological comparisons beyond the planned hypothesis tests are exploratory in nature. Generalizability to other samples can thus not be directly concluded.

Data used here were acquired at multiple sites. It is crucial for actual clinical implementations that models generalize to different sites, ideally without a need for site-specific training datasets to address potential between-scanner differences (88, 89). Though MVPA techniques ideally separate informative from uninformative information in the data, it cannot be excluded that site effects may have reduced diagnostic performance in this study.

Models reported in this paper were limited to SVM and limited exploratory assessment of GPC classifiers. Thus, despite similar diagnostic performance of SVM and GPC, results do not generalize to other MVPA algorithms available for analyses of neuroimaging data (2, 6, 7).

CONCLUSION

Support vector machine-classification of fMRI data from an interoception task did not prove to be diagnostically applicable to predict individual CBT outcomes in PD/AG as measured by general anxiety reduction. Results contrast with previously reported diagnostic accuracies in models based on alternative tasks with alternative classifiers. Results of an exploratory analysis indicate that the method may be more suitable to predict symptom reduction directly related to an IE and related extent of interoceptive symptom relief. We believe that beyond identifying optimal data analysis strategies, the identification and optimization of suitable paradigms will be an important area of research in developing neuroimaging biomarkers in PD/AG and other anxiety disorders.

ETHICS STATEMENT

Written informed consent was obtained from all participants in accordance with the Declaration of Helsinki. The randomized

clinical trial (isrctn.org identifier: ISRCTN80046034) was approved by the ethics committee of the Medical Faculty of the Technische Universität Dresden (agreement EK 164082006). The neuroimaging components were approved by the ethics committee of the Medical Faculty of the RWTH Aachen University, Aachen (agreement EK 073/07) and at all local sites.

AUTHOR CONTRIBUTIONS

BS, JB, and BP planned and carried out the analyses and drafted the manuscript. UL, DW, AG, BStr, H-UW, AS, AW, CK, TK, VA, and BP planned and carried out data acquisition for the MAC trial and neuroimaging substudy. They contributed to the interpretation of the results and to revision of the manuscript for important intellectual content. All the authors approved the final version of the manuscript and agreed to be accountable for the content of the work.

ACKNOWLEDGMENTS

The authors thank all participants of this study and the entire team of the PANIC-NET involved in the MAC study. For a comprehensive list of people involved see Ref. (33). We acknowledge support by the Open Access Publication Fund of the University of Muenster. Special thanks go to Dominik Grotegerd (University Hospital Münster, Department of Psychiatry) for support with the MANIA Toolbox and to Prof. Ulrich Stöber (University of Applied Sciences, Münster) for co-supervising JB's bachelor thesis based on this project. Results of preliminary analyses have been presented in the form of abstracts at the annual meetings of the German Society of Neuroradiology (90) and the German Chapter of the International Society for Magnetic Resonance in Medicine in 2015.

FUNDING

The MAC study was funded by the German Federal Ministry of Education and Research (01GV0615) as part of the Psychotherapy Research Funding Initiative.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fpsy.2017.00099/full#supplementary-material>.

REFERENCES

1. Abi-Dargham A, Horga G. The search for imaging biomarkers in psychiatric disorders. *Nat Med* (2016) 22:1248–55. doi:10.1038/nm.4190
2. Arbabshirani MR, Plis S, Sui J, Calhoun VD. Single subject prediction of brain disorders in neuroimaging: promises and pitfalls. *Neuroimage* (2017) 145:137–65. doi:10.1016/j.neuroimage.2016.02.079
3. Haller S, Lovblad KO, Giannakopoulos P, Van De Ville D. Multivariate pattern recognition for diagnosis and prognosis in clinical neuroimaging: state of the art, current challenges and future trends. *Brain Topogr* (2014) 27:329–37. doi:10.1007/s10548-014-0360-z
4. Klöppel S, Abdulkadir A, Jack CR Jr, Koutsouleris N, Mourao-Miranda J, Vemuri P. Diagnostic neuroimaging across diseases. *Neuroimage* (2012) 61(2):457–63. doi:10.1016/j.neuroimage.2011.11.002
5. Orru G, Petterson-Yeo W, Marquand AF, Sartori G, Mechelli A. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci Biobehav Rev* (2012) 36:1140–52. doi:10.1016/j.neubiorev.2012.01.004
6. Sundermann B, Herr D, Schwindt W, Pfeleiderer B. Multivariate classification of blood oxygen level-dependent fMRI data with diagnostic intention: a clinical perspective. *AJNR Am J Neuroradiol* (2014) 35:848–55. doi:10.3174/ajnr.A3713

7. Wolfers T, Buitelaar JK, Beckmann CF, Franke B, Marquand AF. From estimating activation locality to predicting disorder: a review of pattern recognition for neuroimaging-based psychiatric diagnostics. *Neurosci Biobehav Rev* (2015) 57:328–49. doi:10.1016/j.neubiorev.2015.08.001
8. Woo CW, Chang LJ, Lindquist MA, Wager TD. Building better biomarkers: brain models in translational neuroimaging. *Nat Neurosci* (2017) 20:365–77. doi:10.1038/nn.4478
9. Ball TM, Stein MB, Paulus MP. Toward the application of functional neuroimaging to individualized treatment for anxiety and depression. *Depress Anxiety* (2014) 31:920–33. doi:10.1002/da.22299
10. Doehrmann O, Ghosh SS, Polli FE, Reynolds GO, Horn F, Keshavan A, et al. Predicting treatment response in social anxiety disorder from functional magnetic resonance imaging. *JAMA Psychiatry* (2013) 70:87–97. doi:10.1001/2013.jamapsychiatry.5
11. Hahn T, Kircher T, Straube B, Wittchen HU, Konrad C, Strohle A, et al. Predicting treatment response to cognitive behavioral therapy in panic disorder with agoraphobia by integrating local neural information. *JAMA Psychiatry* (2015) 72(1):68–74. doi:10.1001/jamapsychiatry.2014.1741
12. Lueken U, Zierhut KC, Hahn T, Straube B, Kircher T, Reif A, et al. Neurobiological markers predicting treatment response in anxiety disorders: a systematic review and implications for clinical application. *Neurosci Biobehav Rev* (2016) 66:143–62. doi:10.1016/j.neubiorev.2016.04.005
13. Mansson KN, Frick A, Boraxbekk CJ, Marquand AF, Williams SC, Carlbring P, et al. Predicting long-term outcome of Internet-delivered cognitive behavior therapy for social anxiety disorder using fMRI and support vector machine learning. *Transl Psychiatry* (2015) 5:e530. doi:10.1038/tp.2015.22
14. Phillips ML, Chase HW, Sheline YI, Etkin A, Almeida JR, Deckersbach T, et al. Identifying predictors, moderators, and mediators of antidepressant response in major depressive disorder: neuroimaging approaches. *Am J Psychiatry* (2015) 172:124–38. doi:10.1176/appi.ajp.2014.14010076
15. Whitfield-Gabrieli S, Ghosh SS, Nieto-Castanon A, Saygin Z, Doehrmann O, Chai XJ, et al. Brain connectomics predict response to treatment in social anxiety disorder. *Mol Psychiatry* (2016) 21:680–5. doi:10.1038/mp.2015.109
16. Linden DE. How psychotherapy changes the brain – the contribution of functional neuroimaging. *Mol Psychiatry* (2006) 11:528–38. doi:10.1038/sj.mp.4001816
17. Lueken U, Hahn T. Functional neuroimaging of psychotherapeutic processes in anxiety and depression: from mechanisms to predictions. *Curr Opin Psychiatry* (2016) 29:25–31. doi:10.1097/YCO.0000000000000218
18. Porto PR, Oliveira L, Mari J, Volchan E, Figueira I, Ventura P. Does cognitive behavioral therapy change the brain? A systematic review of neuroimaging in anxiety disorders. *J Neuropsychiatry Clin Neurosci* (2009) 21:114–25. doi:10.1176/jnp.2009.21.2.114
19. Roffman JL, Marci CD, Glick DM, Dougherty DD, Rauch SL. Neuroimaging and the functional neuroanatomy of psychotherapy. *Psychol Med* (2005) 35:1385–98. doi:10.1017/S0033291705005064
20. Serretti A, Chiesa A, Calati R, Perna G, Bellodi L, De Ronchi D. Common genetic, clinical, demographic and psychosocial predictors of response to pharmacotherapy in mood and anxiety disorders. *Int Clin Psychopharmacol* (2009) 24:1–18. doi:10.1097/YIC.0b013e32831db2d7
21. Yang Y, Kircher T, Straube B. The neural correlates of cognitive behavioral therapy: recent progress in the investigation of patients with panic disorder. *Behav Res Ther* (2014) 62:88–96. doi:10.1016/j.brat.2014.07.011
22. Lener MS, Iosifescu DV. In pursuit of neuroimaging biomarkers to guide treatment selection in major depressive disorder: a review of the literature. *Ann N Y Acad Sci* (2015) 1344:50–65. doi:10.1111/nyas.12759
23. Bandelow B, Lichte T, Rudolf S, Wiltink J, Beutel M. *S3-Leitlinie Angststörungen*. Berlin, Heidelberg: Springer (2015).
24. Furukawa TA, Watanabe N, Churchill R. Combined psychotherapy plus antidepressants for panic disorder with or without agoraphobia. *Cochrane Database Syst Rev* (2007) (1):CD004364. doi:10.1002/14651858.CD004364.pub2
25. McHugh RK, Smits JA, Otto MW. Empirically supported treatments for panic disorder. *Psychiatr Clin North Am* (2009) 32:593–610. doi:10.1016/j.psc.2009.05.005
26. Mitte K. A meta-analysis of the efficacy of psycho- and pharmacotherapy in panic disorder with and without agoraphobia. *J Affect Disord* (2005) 88:27–45. doi:10.1016/j.jad.2005.05.003
27. Otto MW, Deveney C. Cognitive-behavioral therapy and the treatment of panic disorder: efficacy and strategies. *J Clin Psychiatry* (2005) 66(Suppl 4):28–32.
28. Roy-Byrne PP, Craske MG, Stein MB. Panic disorder. *Lancet* (2006) 368:1023–32. doi:10.1016/S0140-6736(06)69418-X
29. Starcevic V. Treatment of panic disorder: recent developments and current status. *Expert Rev Neurother* (2008) 8:1219–32. doi:10.1586/14737175.8.8.1219
30. Hofmann SG, Asnaani A, Vonk IJ, Sawyer AT, Fang A. The efficacy of cognitive behavioral therapy: a review of meta-analyses. *Cognit Ther Res* (2012) 36:427–40. doi:10.1007/s10608-012-9476-1
31. Kelly JM, Jakubovski E, Bloch MH. Prognostic subgroups for remission and response in the Coordinated Anxiety Learning and Management (CALM) trial. *J Clin Psychiatry* (2015) 76:267–78. doi:10.4088/JCP.13m08922
32. Schneider RL, Arch JJ, Wolitzky-Taylor KB. The state of personalized treatment for anxiety disorders: a systematic review of treatment moderators. *Clin Psychol Rev* (2015) 38:39–54. doi:10.1016/j.cpr.2015.02.004
33. Kircher T, Arolt V, Jansen A, Pyka M, Reinhardt I, Kellermann T, et al. Effect of cognitive-behavioral therapy on neural correlates of fear conditioning in panic disorder. *Biol Psychiatry* (2013) 73:93–101. doi:10.1016/j.biopsych.2012.07.026
34. Lueken U, Straube B, Konrad C, Wittchen HU, Strohle A, Wittmann A, et al. Neural substrates of treatment response to cognitive-behavioral therapy in panic disorder with agoraphobia. *Am J Psychiatry* (2013) 170:1345–55. doi:10.1176/appi.ajp.2013.12111484
35. Lueken U, Straube B, Wittchen HU, Konrad C, Strohle A, Wittmann A, et al. Therapygenetics: anterior cingulate cortex-amygdala coupling is associated with 5-HTTLPR and treatment response in panic disorder with agoraphobia. *J Neural Transm* (2015) 122:135–44. doi:10.1007/s00702-014-1311-2
36. Straube B, Lueken U, Jansen A, Konrad C, Gloster AT, Gerlach AL, et al. Neural correlates of procedural variants in cognitive-behavioral therapy: a randomized, controlled multicenter fMRI study. *Psychother Psychosom* (2014) 83:222–33. doi:10.1159/000359955
37. Ball TM, Stein MB, Ramsawh HJ, Campbell-Sills L, Paulus MP. Single-subject anxiety treatment outcome prediction using functional neuroimaging. *Neuropsychopharmacology* (2014) 39:1254–61. doi:10.1038/npp.2013.328
38. Bouton ME, Mineka S, Barlow DH. A modern learning theory perspective on the etiology of panic disorder. *Psychol Rev* (2001) 108:4–32. doi:10.1037/0033-295X.108.1.4
39. Domschke K, Stevens S, Pfleiderer B, Gerlach AL. Interoceptive sensitivity in anxiety and anxiety disorders: an overview and integration of neurobiological findings. *Clin Psychol Rev* (2010) 30:1–11. doi:10.1016/j.cpr.2009.08.008
40. Kleint NI, Wittchen HU, Lueken U. Probing the interoceptive network by listening to heartbeats: an fMRI study. *PLoS One* (2015) 10:e0133164. doi:10.1371/journal.pone.0133164
41. Pfleiderer B, Berse T, Stroux D, Ewert A, Konrad C, Gerlach AL. Internal focus of attention in anxiety-sensitive females up-regulates amygdala activity: an fMRI study. *J Neural Transm* (2014) 121:1417–28. doi:10.1007/s00702-014-1248-5
42. Schulz SM. Neural correlates of heart-focused interoception: a functional magnetic resonance imaging meta-analysis. *Philos Trans R Soc Lond B Biol Sci* (2016) 371:20160018. doi:10.1098/rstb.2016.0018
43. Gloster AT, Wittchen HU, Einsle F, Hofler M, Lang T, Helbig-Lang S, et al. Mechanism of action in CBT (MAC): methods of a multi-center randomized controlled trial in 369 patients with panic disorder and agoraphobia. *Eur Arch Psychiatry Clin Neurosci* (2009) 259(Suppl 2):S155–66. doi:10.1007/s00406-009-0065-6
44. Arolt V, Zwanzger P, Strohle A, Hamm A, Gerlach A, Kircher T, et al. The research network PANIC-NET: improving the treatment of panic disorder – from a better understanding of fear circuit mechanisms to more effective psychological treatment and routine care. *Psychother Psychosom Med Psychol* (2009) 59:124–31. doi:10.1055/s-0029-1202781
45. American Psychiatric Association and American Psychiatric Association Task Force on DSM-IV. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR*. Washington, DC: American Psychiatric Association (2000).
46. Hamilton M. The assessment of anxiety states by rating. *Br J Med Psychol* (1959) 32:50–5. doi:10.1111/j.2044-8341.1959.tb00467.x
47. Shear MK, Vander Bilt J, Rucci P, Endicott J, Lydiard B, Otto MW, et al. Reliability and validity of a structured interview guide for the Hamilton Anxiety Rating Scale (SIGH-A). *Depress Anxiety* (2001) 13:166–78. doi:10.1002/da.1033.abc

48. Beck AT, Steer RA, Brown GK. *Manual for the BDI-II*. San Antonio: The Psychological Corporation (1996).
49. Reiss S, Peterson RA, Gursky DM, McNally RJ. Anxiety sensitivity, anxiety frequency and the prediction of fearfulness. *Behav Res Ther* (1986) 24:1–8. doi:10.1016/0005-7967(86)90143-9
50. Guy W. Clinical Global Impression. In: Guy W, editor. *ECDEU Assessment Manual for Psychopharmacology, Revised*. Rockville: National Institute of Mental Health (1976). p. 217–22.
51. Barlow DH, Gorman JM, Shear MK, Woods SW. Cognitive-behavioral therapy, imipramine, or their combination for panic disorder: a randomized controlled trial. *JAMA* (2000) 283:2529–36. doi:10.1001/jama.283.19.2529
52. Reitan RM. Trail making test results for normal and brain-damaged children. *Percept Mot Skills* (1971) 33:575–81. doi:10.2466/pms.1971.33.2.575
53. Lepach AC, Daseking M, Petermann F, Waldmann HC. The relationships of intelligence and memory assessed using the WAIS-IV and the WMS-IV. *Gesundheitswesen* (2013) 75:775–81. doi:10.1055/s-0033-1357168
54. Westphal D, Gerlach AL, Lang T, Wittchen H, Hamm AO, Ströhle A, et al. Die Effekte interozeptiver Expositionsübungen in der Kognitiven Verhaltenstherapie von Panikstörung mit Agoraphobie. *Verhaltenstherapie* (2015) 25:268–76. doi:10.1159/000441508
55. Schandry R. Heart beat perception and emotional experience. *Psychophysiology* (1981) 18:483–8. doi:10.1111/j.1469-8986.1981.tb02486.x
56. Grotegerd D, Redlich R, Almeida JR, Riemenschneider M, Kugel H, Arolt V, et al. MANIA – a pattern classification toolbox for neuroimaging data. *Neuroinformatics* (2014) 12:471–86. doi:10.1007/s12021-014-9223-8
57. Pereira F, Mitchell T, Botvinick M. Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage* (2009) 45:S199–209. doi:10.1016/j.neuroimage.2008.11.007
58. Golland P, Fischl B. Permutation tests for classification: towards statistical significance in image-based studies. In: Taylor C, Noble JA, editors. *Information Processing in Medical Imaging. IPMI 2003. Lecture Notes in Computer Science*. Vol 2732. Heidelberg: Springer Berlin (2003). p. 330–41.
59. Chang C, Lin C. LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* (2011) 2, 27(1–27):27.
60. Alpaydin E. *Introduction to Machine Learning*. Cambridge, MA: The MIT Press (2010).
61. James G, Witten D, Hastie T. *An Introduction to Statistical Learning with Applications in R*. New York, Heidelberg, Dordrecht, London: Springer (2013).
62. Vapnik VN. *The Nature of Statistical Learning Theory*. 2nd ed. New York: Springer-Verlag (2000).
63. Mwangi B, Tian TS, Soares JC. A review of feature reduction techniques in neuroimaging. *Neuroinformatics* (2014) 12(2):229–44. doi:10.1007/s12021-013-9204-3
64. Schrouff J, Rosa MJ, Rondina JM, Marquand AF, Chu C, Ashburner J, et al. PRoNT: pattern recognition for neuroimaging toolbox. *Neuroinformatics* (2013) 11:319–37. doi:10.1007/s12021-013-9178-1
65. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Mach Learn* (2002) 46:389–422. doi:10.1023/A:1012487302797
66. Ding Y, Wilkins D. Improving the performance of SVM-RFE to select genes in microarray data. *BMC Bioinformatics* (2006) 7(Suppl 2):S12. doi:10.1186/1471-2105-7-S2-S12
67. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated Anatomical Labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* (2002) 15:273–89. doi:10.1006/nimg.2001.0978
68. Desikan RS, Segonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* (2006) 31:968–80. doi:10.1016/j.neuroimage.2006.01.021
69. Frazier JA, Chiu S, Breeze JL, Makris N, Lange N, Kennedy DN, et al. Structural brain magnetic resonance imaging of limbic and thalamic volumes in pediatric bipolar disorder. *Am J Psychiatry* (2005) 162:1256–65. doi:10.1176/appi.ajp.162.7.1256
70. Goldstein JM, Seidman LJ, Makris N, Ahern T, O'Brien LM, Caviness VS Jr, et al. Hypothalamic abnormalities in schizophrenia: sex effects and genetic vulnerability. *Biol Psychiatry* (2007) 61:935–45. doi:10.1016/j.biopsych.2006.06.027
71. Makris N, Goldstein JM, Kennedy D, Hodge SM, Caviness VS, Faraone SV, et al. Decreased volume of left and total anterior insular lobule in schizophrenia. *Schizophr Res* (2006) 83:155–71. doi:10.1016/j.schres.2005.11.020
72. Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, et al. Automated Talairach atlas labels for functional brain mapping. *Hum Brain Mapp* (2000) 10:120–31. doi:10.1002/1097-0193(200007)10:3<120::AID-HBM30>3.0.CO;2-8
73. Talairach JJ, Tournoux P. *Co-Planar Stereotaxic Atlas of the Human Brain*. New York: Thieme Medical Publishers (1988).
74. Rasmussen CE, Nickisch H. Gaussian Processes for Machine Learning (GPML) toolbox. *J Mach Learn Res* (2010) 11:3011–5.
75. Slotnick SD, Moo LR, Segal JB, Hart J Jr. Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Brain Res Cogn Brain Res* (2003) 17:75–82. doi:10.1016/S0926-6410(03)00082-X
76. Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci* (2006) 10:424–30. doi:10.1016/j.tics.2006.07.005
77. Lueken U, Straube B, Yang Y, Hahn T, Beesdo-Baum K, Wittchen HU, et al. Separating depressive comorbidity from panic disorder: a combined functional magnetic resonance imaging and machine learning approach. *J Affect Disord* (2015) 184:182–92. doi:10.1016/j.jad.2015.05.052
78. Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* (2013) 14:365–76. doi:10.1038/nrn3475
79. Ioannidis JP. Why most published research findings are false. *PLoS Med* (2005) 2:e124. doi:10.1371/journal.pmed.0020124
80. Khalsa SS, Lapidus RC. Can interoception improve the pragmatic search for biomarkers in psychiatry? *Front Psychiatry* (2016) 7:121. doi:10.3389/fpsy.2016.00121
81. Chu C, Hsu AL, Chou KH, Bandettini P, Lin C; Alzheimer's Disease, et al. Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *Neuroimage* (2012) 60:59–70. doi:10.1016/j.neuroimage.2011.11.066
82. Clarke R, Ressom HW, Wang A, Xuan J, Liu MC, Gehan EA, et al. The properties of high-dimensional data spaces: implications for exploring gene and protein expression data. *Nat Rev Cancer* (2008) 8:37–49. doi:10.1038/nrc2294
83. Pappu V, Pardalos PM. High-dimensional data classification. In: Aleskerov F, Goldengorin B, Pardalos PM, editors. *Clusters, Orders, and Trees: Methods and Applications: In Honor of Boris Mirkin's 70th Birthday*. New York, NY: Springer (2014). p. 119–50.
84. Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci* (2009) 12:535–40. doi:10.1038/nn.2303
85. Poline JB, Kherif F, Penny W. Contrasts and classical inference. In: Frackowiak RSJ, Friston KJ, Frith CD, Dolan RJ, Price CJ, Zeki S, et al., editors. *Human Brain Function*. 2nd ed. Chapter 38. Burlington: Academic Press (2004). p. 761–79.
86. Drucker H, Burges CJ, Kaufman L, Smola A, Vapnik V. Support vector regression machines. In: Mozer MC, Jordan JI, Petsche T, editors. *Advances in Neural Information Processing Systems 9*. Cambridge: MIT Press (1997). p. 155–61.
87. Furnham A. Response bias, social desirability and dissimulation. *Pers Individ Dif* (1986) 7:385–400. doi:10.1016/0191-8869(86)90014-0
88. Friedman L, Glover GH, Fbrn Consortium. Reducing interscanner variability of activation in a multicenter fMRI study: controlling for signal-to-fluctuation-noise-ratio (SFNR) differences. *Neuroimage* (2006) 33:471–81. doi:10.1016/j.neuroimage.2006.07.012
89. Zou KH, Greve DN, Wang M, Pieper SD, Warfield SK, White NS, et al. Reproducibility of functional MR imaging: preliminary results of prospective multi-institutional study performed by Biomedical Informatics Research Network. *Radiology* (2005) 237:781–9. doi:10.1148/radiol.2373041630
90. Sundermann B, Bode J, Gerlach AL, Konrad C, Lueken U, Straube B, et al. Diagnostische Genauigkeit der Interozeptions-fMRT zur Prädiktion des Behandlungserfolgs einer kognitiven Verhaltenstherapie bei Patienten mit Panikstörung und Agoraphobie (PD/AG). *Clin Neuroradiol* (2015) 25 (Suppl 1):0106. doi:10.1007/s00062-015-0445-4

Conflict of Interest Statement: VA is member of the advisory boards and/or gave presentations for the following companies: Astra-Zeneca, Janssen-Organon, Lilly, Lundbeck, Servier, Pfizer, and Wyeth. He chaired the committee for the Wyeth Research Award Depression and Anxiety. TK has received

in the past 3 years honoraria or educational grants from Janssen, Bristol Myers-Squibb, Wyeth, Lundbeck, Lilly, Astra-Zeneca, and Pfizer. CK received fees for an educational program from Aristo Pharma, Janssen-Cilag, Lilly, MagVenture, Servier, and Trommsdorff as well as travel support and speakers honoraria from Aristo Pharma, Janssen-Cilag, Lundbeck and Servier. The other authors reported no biomedical financial interests or potential conflicts of interest.

Copyright © 2017 Sundermann, Bode, Lueken, Westphal, Gerlach, Straube, Wittchen, Ströhle, Wittmann, Konrad, Kircher, Arolt and Pfeleiderer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.