# Detecting Neuroimaging Biomarkers for Schizophrenia: A Meta-Analysis of Multivariate Pattern Recognition Studies

**Joseph Kambeitz\*[1,2], Lana Kambeitz-Ilankovic[1], Stefan Leucht[3], Stephen Wood[4], Christos Davatzikos[5], Berend Malchow[1], Peter Falkai[1] and Nikolaos Koutsouleris[1]**

[1]Department of Psychiatry, Ludwig-Maximilians University Munich, Munich, Germany; [2]King's College London, Psychosis Clinical Academic Group, Institute of Psychiatry, London, UK; [3]Department of Psychiatry, Technical University Munich, Munich, Germany; [4]School of Psychology, University of Birmingham, Birmingham, UK; [5]Section of Biomedical Image Analysis, Department of Radiology, University of Pennsylvania, Pennsylvania, USA

Multivariate pattern recognition approaches have recently facilitated the search for reliable neuroimaging-based biomarkers in psychiatric disorders such as schizophrenia. By taking into account the multivariate nature of brain functional and structural changes as well as their distributed localization across the whole brain, they overcome drawbacks of traditional univariate approaches. To evaluate the overall reliability of neuroimaging-based biomarkers, we conducted a comprehensive literature search to identify all studies that used multivariate pattern recognition to identify patterns of brain alterations that differentiate patients with schizophrenia from healthy controls. A bivariate random-effects meta-analytic model was implemented to investigate the sensitivity and specificity across studies as well as to assess the robustness to potentially confounding variables. In the total sample of $n = 38$ studies (1602 patients and 1637 healthy controls), patients were differentiated from controls with a sensitivity of 80.3% (95% CI: 76.7–83.5%) and a specificity of 80.3% (95% CI: 76.9–83.3%). Analysis of neuroimaging modality indicated higher sensitivity (84.46%, 95% CI: 79.9–88.2%) and similar specificity (76.9%, 95% CI: 71.3–81.6%) of rsfMRI studies as compared with structural MRI studies (sensitivity: 76.4%, 95% CI: 71.9–80.4%, specificity of 79.0%, 95% CI: 74.6–82.8%). Moderator analysis identified significant effects of age ($p = 0.029$), imaging modality ($p = 0.019$), and disease stage ($p = 0.025$) on sensitivity as well as of positive-to-negative symptom ratio ($p = 0.022$) and antipsychotic medication ($p = 0.016$) on specificity. Our results underline the utility of multivariate pattern recognition approaches for the identification of reliable neuroimaging-based biomarkers. Despite the clinical heterogeneity of the schizophrenia phenotype, brain functional and structural alterations differentiate schizophrenic patients from healthy controls with 80% sensitivity and specificity.

*Neuropsychopharmacology* (2015) **40,** 1742–1751; doi:10.1038/npp.2015.22; published online 18 February 2015

## INTRODUCTION

Schizophrenia shows a lifetime prevalence of 0.30–0.66% in the general population(McGrath *et al*, 2008) making it one of the leading factors of global disease burden ('WHO, 2004 Global Burden of Disease—Update,' n.d.). After more than a century of research into the neurobiology of the disorder, its pathophysiological underpinnings still remain unknown. Over the past 15 years, considerable research efforts elucidated a vast array of functional (Howes *et al*, 2012; Minzenberg *et al*, 2009) and structural brain abnormalities (Chan *et al*, 2009; Fornito *et al*, 2009; Honea *et al*, 2005) that may constitute the 'organic surrogate' of the illness. Even though these results indicate significant differences in, eg, brain structure between healthy controls (HC) and patients

at the group level—a substantial overlap is usually observed between groups, which disallow the use of these differences for the individualized diagnosis of the disorder. Therefore, alterations in brain structure and function have so far not been successfully integrated into the diagnostic process as disease biomarkers operating on the single-subject level (Borgwardt *et al*, 2012; Borgwardt and Fusar-Poli, 2012; Kapur *et al*, 2012). The main reason for this gap between research and its potential diagnostic application lies in traditional univariate statistical approaches, which neglect the heavily interconnected nature of the functional and structural brain data (Davatzikos, 2004).

To overcome these methodological drawbacks, an increasing number of studies have applied novel multivariate statistical approaches to the analysis of brain alterations in patients with schizophrenia (eg, (Davatzikos *et al*, 2005; Fu and Costafreda, 2013; McIntosh and Lobaugh, 2004; Zarogianni *et al*, 2013)). These results indicate that patterns of subtle structural and functional changes can be highly distinctive of schizophrenia-related brain alterations, even though each individual component within these patterns might be not. Most importantly, the classification perfor-

*Correspondence: J Kambeitz, Department of Psychiatry, University of Munich, Nußbaumstraße 7, 80336 Munich, Germany, Tel: +498 944 005 5700, Fax: +498 944 005 5708,
E-mail: joseph.kambeitz@med.uni-muenchen.de

mance of neuroimaging biomarkers based on multivariate statistical methods is typically assessed by using cross-validation strategies that allow estimating the predictive models' generalizability to unseen test individuals. In this regard, the majority of studies using multivariate machine learning algorithms reported good generalization performances, which might open up the possibility of neuroimaging to become part of the routine diagnostic process in the future. For instance, support-vector machines (Davatzikos et al, 2005), partial least squares analysis (Kawasaki et al, 2007; McIntosh and Lobaugh, 2004), random forests (Anderson et al, 2010; Greenstein et al, 2012) and artificial neural networks (Bose et al, 2008; Josin and Liddle, 2001; Rathi et al, 2010) have shown to differentiate patients from HC with diagnostic accuracies of 60–100% using neuroimaging data.

However, these studies differ with respect to multiple aspects such as the demographic characteristics of the investigated populations, the clinical symptoms of the patient samples, the imaging modalities employed, the preprocessing of neuroimaging data prior to analysis, the statistical models, as well as the evaluation scheme of the models' performance. As a result, the sensitivity and specificity of the reported predictive models differ widely, making it difficult to compare the classification performance of neuroimaging-based biomarkers across studies. Furthermore, little is known about which factors contribute to the success of MRI-based predictive modeling as authors may typically test a range of analysis pipelines and finally report only the analysis scheme achieving the highest test performance (Pers et al, 2009). Only a few studies have systematically compared two or more algorithms (Bose et al, 2008; Castellani et al, 2012; Rathi et al, 2010). However, a systematic investigation of different imaging modalities or multivariate methods is still missing. Finally, to the best of our knowledge, no comparative reports exist to date on the relationship between clinical variables of the tested samples and diagnostic accuracies of neuroimaging-based diagnostic models. Age, gender, psychiatric symptoms, or current medication represent potentially confounding variables, which might affect the diagnostic success of such models.

Thus, we conducted a meta-analysis of multivariate pattern recognition studies to evaluate the performance of neuroimaging phenotypes in distinguishing patients with schizophrenia from HC. Within this framework, we also assessed the potentially moderating impact of different clinical variables on these neurodiagnostic signatures.

## MATERIALS AND METHODS

### Search and Selection Strategy

The entire electronic PubMed database was searched from 1 January 1950 up to 31 May 2013. Initially, studies were screened by using a comprehensive search term (('support vector' OR 'SVM' OR 'classification' OR 'categorization') AND ('MRI' OR 'fMRI' OR 'magnetic resonance' OR 'imaging' OR 'gray matter' OR 'gray matter' OR 'white matter' OR 'DTI' OR 'diffusion tensor imaging' OR 'PET' OR 'positron emission tomography' OR 'SPECT' OR 'single photon emission tomography') AND ('schizophrenia' OR 'psychosis' OR 'psychotic' OR 'schizophreniform')).

Subsequently, all studies were screened according to the following criteria: To be included in the meta-analysis a paper needed to report results of a neuroimaging-based multivariate classification model separating patients with schizophrenia from HC. We included all available multivariate approaches such as support-vector machines, random forests, discriminant analysis, logistic regression, neural networks, as well as combinations thereof. Studies were included if the following measures of classification performance were available or if data allowed for the calculation of the following parameters: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). In case insufficient data was reported, authors were contacted via email to provide additional information regarding their published reports. Furthermore, in multivariate classification it is of utmost importance to apply some form of cross-validation while estimating model parameters to avoid overfitting, which is associated with low generalizability. Thus, only studies that applied cross-validation (eg, leave-one-out, n-fold, and bootstrapping) were included in the analysis. In some cases, multiple studies were published based on the same sample or with large overlap between samples. We verified sample overlap by contacting the corresponding authors. In order to avoid bias we excluded samples with large overlap (shared $n > 20\%$). The results of the literature search are presented in a flow-chart following the PRISMA guidelines (Moher et al, 2009) (see Supplementary Figure 1).

### Data Extraction

The main outcome measure was the diagnostic test performance of the different multivariate approaches for separating schizophrenic patients from HC as measured by sensitivity $(= TP/(TP + FN))$ and specificity $(= TN/(TN + FP))$. The following additional information was extracted from all studies: names of the authors; year of publication; population characteristics of HC and patient groups (group size, age, gender, antipsychotic use, diagnosis, and symptom ratings); type of neuroimaging data (magnetic resonance imaging 'MRI', functional MRI 'fMRI', resting-state fMRI 'rsfMRI', positron emission tomography 'PET', single photon emission computed tomography 'SPECT', diffusion tensor imaging 'DTI', scanner type, and resolution), characteristics of the employed preprocessing methodology, characteristics of the classification method (eg, linear discriminant analysis and support-vector machine) and characteristics of the cross-validation procedure. Data extraction was performed by two authors separately (LKI, JK) to ensure accuracy and disagreements were discussed in a consensus conference. The QUADAS-2 guidelines were used to assess study quality of all publication included in the present meta-analysis (see Supplementary Figure 2) (Whiting et al, 2011).

### Data Analysis

In studies of diagnostic test accuracy sensitivity and specificity are often negatively correlated and therefore pooling them in the context of a meta-analysis might lead to biased results (Gatsonis and Paliwal, 2006). Instead a bivariate approach(Reitsma et al, 2005) and a strategy based on a

hierarchical summary ROC model (HSROC (Rutter and Gatsonis, 2001)) have been suggested to estimate sensitivity and specificity across studies. However, in most situations both approaches lead to identical results (Harbord *et al*, 2007). In the present analysis we implemented the strategy introduced by Reitsma *et al* (2005). In this bivariate approach, log-transformed sensitivity and specificity are combined in one bivariate regression model while explicitly accounting for their correlation. It is assumed that sensitivity and specificity vary across studies because of differences in study populations, sampling errors, and differences in implicit thresholds applied to the data to separate patients from HC. Thus a random-effects model is applied in order to account for between-study heterogeneity. As larger samples are associated with smaller sampling error and thus with more precise effect size estimates, the studies included in the meta-analysis are weighted according to their sample size. Meta-analysis results are presented in forest plots separately for sensitivity and specificity. Summary estimates for sensitivity and specificity are provided separately for MRI, for rsfMRI studies as well as for all studies combined. We considered $n = 5$ to be the minimum number of studies to justify a separate meta-analysis (Ioannidis and Lau, 2001). The robustness of the results as well as the effect of potentially confounding variables (eg, age, gender ratio, and year of publication) was investigated by adding moderator variables to the bivariate regression model. In order to investigate potential publication bias in meta-analyses of diagnostic accuracies, it has been recommended to create funnel plots by plotting log diagnostic odds ratios (logDOR) for all studies against $\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ with $n_1$ and $n_2$ representing the sample sizes of the patient and the HC group. This measure is proportional to the inverted square-root of the effective sample size (ESS):. In case of a publication bias the distribution of studies in the funnel plot is asymmetrical. A statistical test for funnel plot asymmetry is provided by a regression of logDOR with $\frac{1}{\sqrt{ESS}}$ weighted by ESS (Deeks *et al*, 2005). All computations were performed using the R statistical programming language version 2.10.13 (R Core Team, 2013) with the package mada (Doebler, 2012).

## RESULTS

The initial literature search identified 399 studies of interest. After screening all studies and applying the inclusion criteria, 361 studies were excluded. Fan *et al* (Fan *et al*, 2007) and Davatzikos *et al* (Davatzikos *et al*, 2005) used overlapping samples. Only Fan *et al* (Fan *et al*, 2007) was included in the main analysis as it is the most recent report of this sample. For additional moderator analysis we included Davatzikos *et al* as additional data was provided (Davatzikos *et al*, 2005). Between Liu *et al* (Liu *et al*, 2012) and Shen *et al* (Shen *et al*, 2010) there was an overlap of only 4 out of 32 subjects. This was considered a minor overlap and both samples were included in the analysis. Two studies (Fekete *et al*, 2013; Hu *et al*, 2013) were based on the same sample but were included in the meta-analysis as they computed predictive models based on fundamentally different features. In order to exclude the possibility

that this affected the results of our meta-analysis, the effect of excluding each of these studies on overall sensitivity and specificity was investigated. The final sample consisted of $n = 38$ studies with of a total of $n = 1602$ SZ patients and $n = 1637$ HC. Among the included studies were $n = 20$ studies using structural MRI, $n = 11$ studies using rsfMRI, $n = 4$ studies using fMRI, $n = 3$ studies using PET, and $n = 1$ study using DTI to build predictive models (see Supplementary Table 1 for an overview of the characteristics of the included studies).

Across all studies, neuroimaging-based classifiers separated SZ from HC with a sensitivity of 80.3% (95% CI: 76.7–83.5%, see Figure 1) and a specificity of 80.3% (95% CI: 76.9–83.3%, see Figure 2). A summary ROC-curve of the included studies along with the estimated summary is presented in Figure 3. Visual inspection of funnel plots did not show evidence for a publication bias (see Supplementary Figure 3). Regression with year of publication did not show any effect on sensitivity ($p = 0.766$) or specificity ($p = 0.801$).

No significant effects of sex, illness duration, PANSS positive scores, PANSS negative scores, or analysis methods (SVM/LDA) on sensitivity or specificity (all $p > 0.1$) were observed (Table 1). We detected a significant effect of patients' age ($p = 0.029$) indicating higher sensitivity in older subjects (see Figure 4). There was no evidence for an effect of age on specificity ($p = 0.095$) and no age effect in the HC on sensitivity ($p = 0.168$) or specificity ($p = 0.380$). We observed a significant effect of positive-to-negative symptom ratio on specificity ($p = 0.022$), indicating higher specificity in patients with predominantly positive symptoms (see Figure 4). There was no effect of positive-to-negative symptom ratio on sensitivity ($p = 0.500$). Comparing studies investigating first-episode patients (FEP) *vs* chronic patients (CSZ), we found a significantly higher sensitivity in CSZ ($p = 0.025$, see Figure 4) but no such effect on specificity ($p = 0.202$). A significant effect of antipsychotic medication (converted to chlorpromazine equivalents, CPZ-eq) on specificity ($p = 0.016$) was found, indicating higher specificity in subjects treated with higher medication doses (see Figure 4). However, CPZ-eq did not significantly affect sensitivity ($p = 0.09$).

When the structural MRI studies were separately analyzed the meta-analysis showed a sensitivity of 76.4% (95% CI: 71.9–80.4%) and a specificity of 79.0% (95% CI: 74.6–82.8%). The rsfMRI studies had a sensitivity of 84.46% (95% CI: 79.9–88.2%) and a specificity of 76.9% (95% CI: 71.3–81.6%). After excluding Hu *et al* (2013) or Fekete *et al* (2013) there was no significant change in the sensivitiy (84.7% with a 95% CI: 79.98–88.46% and 84.28% with a 95% CI: 79.6–88.04%, respectively) or specificity (77.87% with a 95% CI: 83.3–71.29% and 76.5% with a 95% CI: 81.35–70.84%). 'Data source' was added as a moderating variable to the bivariate meta-analysis model to investigate significant differences between different data sources (MRI, rsfMRI). There was a significant difference ($p = 0.010$) between the sensitivity of rsfMRI and structural MRI studies, indicating higher sensitivity in rsfMRI studies (see Figure 4). There was no significant difference in specificity (see Figure 4). To investigate the potential effect of different multivariate approaches, the data set was restricted to studies that applied support-vector machines ($n = 12$) and discriminant analysis ($n = 13$). The bivariate meta-analytic model showed no significant difference between DA and

| Study | Source | | sensitivity [95%–CI] |
|---|---|---|---|
| Zanetti, 2013 | mri | | 0.80 [ 0.69 , 0.88 ] |
| Petterson–Yeo, 2013 | mri | | 0.57 [ 0.36 , 0.76 ] |
| Greenstein, 2012 | mri | | 0.73 [ 0.64 , 0.81 ] |
| Bansal, 2012 | mri | | 0.91 [ 0.79 , 0.97 ] |
| Ota, 2012 | mri | | 0.73 [ 0.53 , 0.87 ] |
| Nieuwenhuis, 2012 | mri | | 0.67 [ 0.61 , 0.73 ] |
| Borgwardt, 2012 | mri | | 0.85 [ 0.65 , 0.94 ] |
| Karageorgiou, 2011 | mri | | 0.64 [ 0.46 , 0.79 ] |
| Takayanagi, 2011 | mri | | 0.84 [ 0.72 , 0.91 ] |
| Santos, 2010 | mri | | 0.69 [ 0.55 , 0.81 ] |
| Rathi, 2010 | mri | | 0.84 [ 0.64 , 0.94 ] |
| Kasparek, 2010 | mri | | 0.66 [ 0.51 , 0.79 ] |
| Sun, 2009 | mri | | 0.85 [ 0.70 , 0.93 ] |
| Fan, 2007 | mri | | 0.95 [ 0.84 , 0.98 ] |
| Fan, 2007 | mri | | 0.81 [ 0.62 , 0.92 ] |
| Kawasaki, 2007 | mri | | 0.76 [ 0.58 , 0.87 ] |
| Yushkevich, 2005 | mri | | 0.71 [ 0.57 , 0.82 ] |
| Csernansky, 2004 | mri | | 0.73 [ 0.59 , 0.83 ] |
| Nakamura, 2004 | mri | | 0.77 [ 0.59 , 0.89 ] |
| Nakamura, 2004 | mri | | 0.82 [ 0.65 , 0.92 ] |
| **RE Model for mri studies** | | | **0.76 [ 0.72 , 0.80 ]** |
| Hu, 2013 | rsfmri | | 0.77 [ 0.48 , 0.93 ] |
| Fekete, 2013 | rsfmri | | 0.94 [ 0.63 , 0.99 ] |
| Guo, 2012 | rsfmri | | 0.86 [ 0.76 , 0.93 ] |
| Bassett, 2012 | rsfmri | | 0.85 [ 0.68 , 0.94 ] |
| Tang, 2012 | rsfmri | | 0.85 [ 0.65 , 0.94 ] |
| Liu, 2012 | rsfmri | | 0.74 [ 0.54 , 0.87 ] |
| Du, 2012 | rsfmri | | 0.91 [ 0.76 , 0.97 ] |
| Fan, 2011 | rsfmri | | 0.89 [ 0.74 , 0.96 ] |
| Anderson, 2010 | rsfmri | | 0.64 [ 0.30 , 0.88 ] |
| Shen, 2010 | rsfmri | | 0.92 [ 0.78 , 0.98 ] |
| Shi, 2007 | rsfmri | | 0.83 [ 0.70 , 0.91 ] |
| **RE Model for rsfmri studies** | | | **0.84 [ 0.80 , 0.88 ]** |
| Yoon, 2012 | fmri | | 0.59 [ 0.45 , 0.71 ] |
| Du, 2012 | fmri | | 0.95 [ 0.80 , 0.99 ] |
| Yang, 2010 | fmri | | 0.79 [ 0.57 , 0.91 ] |
| Costafreda, 2008 | fmri | | 0.89 [ 0.75 , 0.96 ] |
| Josin, 2001 | rCBF–PET | | 0.97 [ 0.77 , 1.00 ] |
| Levy, 1991 | rCBF–PET | | 0.81 [ 0.54 , 0.94 ] |
| Levy, 1991 | rCBF–PET | | 0.73 [ 0.46 , 0.90 ] |
| Bose, 2008 | F–DOPA–PET | | 0.88 [ 0.67 , 0.96 ] |
| Ardekani, 2011 | DTI | | 0.95 [ 0.85 , 0.98 ] |
| **RE Model for all studies combined** | | | **0.80 [ 0.77 , 0.84 ]** |

0.00     0.25     0.50     0.75     1.00
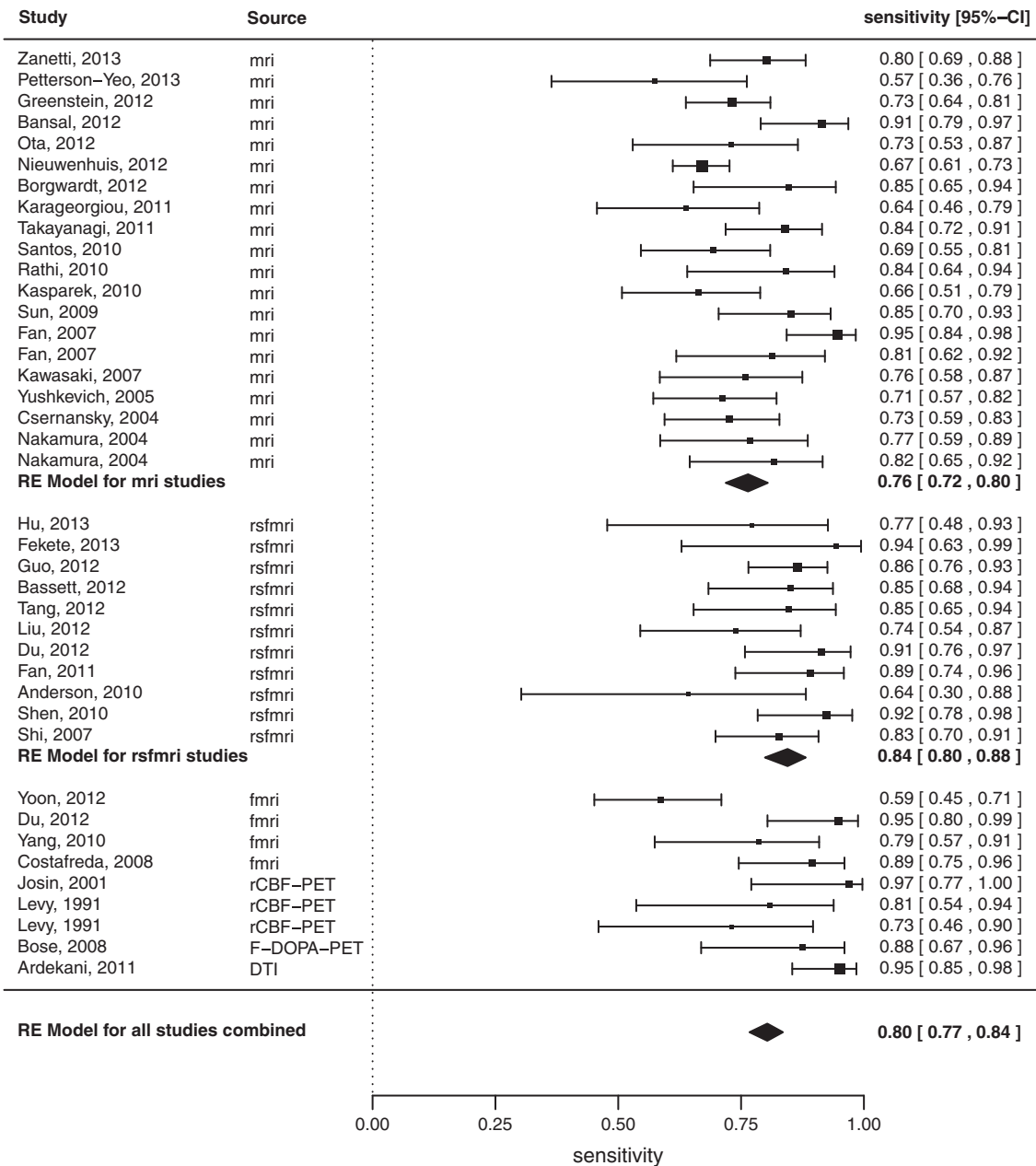
sensitivity

**Figure 1** Forest plot of sensitivities for studies using MRI, fMRI, rsfMRI, rCBF-PET, F-DOPA-PET, and DTI to diagnose schizophrenia. Summary estimates for sensitivity are computed using the approach described by Reitsma *et al* (2005).

SVM studies regarding sensitivity ($p = 0.766$) and specificity ($p = 0.801$).

## DISCUSSION

We present a meta-analysis of a total of $n = 38$ studies with of a total of $n = 1602$ SZ patients and $n = 1637$ HC. Our results suggest that a neuroimaging phenotypes of schizophrenia separate patients from HC with an overall sensitivity and specificity of $\sim 80\%$. Similar results were obtained when the analysis was restricted to individual imaging modalities (structural MRI or rsfMRI). This finding was robust against the inclusion of potential confounding factors such as year of publication and there was no evidence for a publication bias.

## Effect of Age

Interestingly, older age was significantly associated with higher sensitivity. Although illness duration itself did not have a significant impact on sensitivity and specificity, there was a higher sensitivity in patients in a chronic stage of schizophrenia as compared with first-episode patients. These findings might result from more pronounced brain changes in older subjects with schizophrenia. In addition, this finding may be caused by secondary disease effects, which are not related to the underlying brain pathology, but are rather due to environmental factors associated with a more unfavorable illness course in this patient population. In keeping with this hypothesis, numerous studies reported progressive brain changes to be associated with short-term
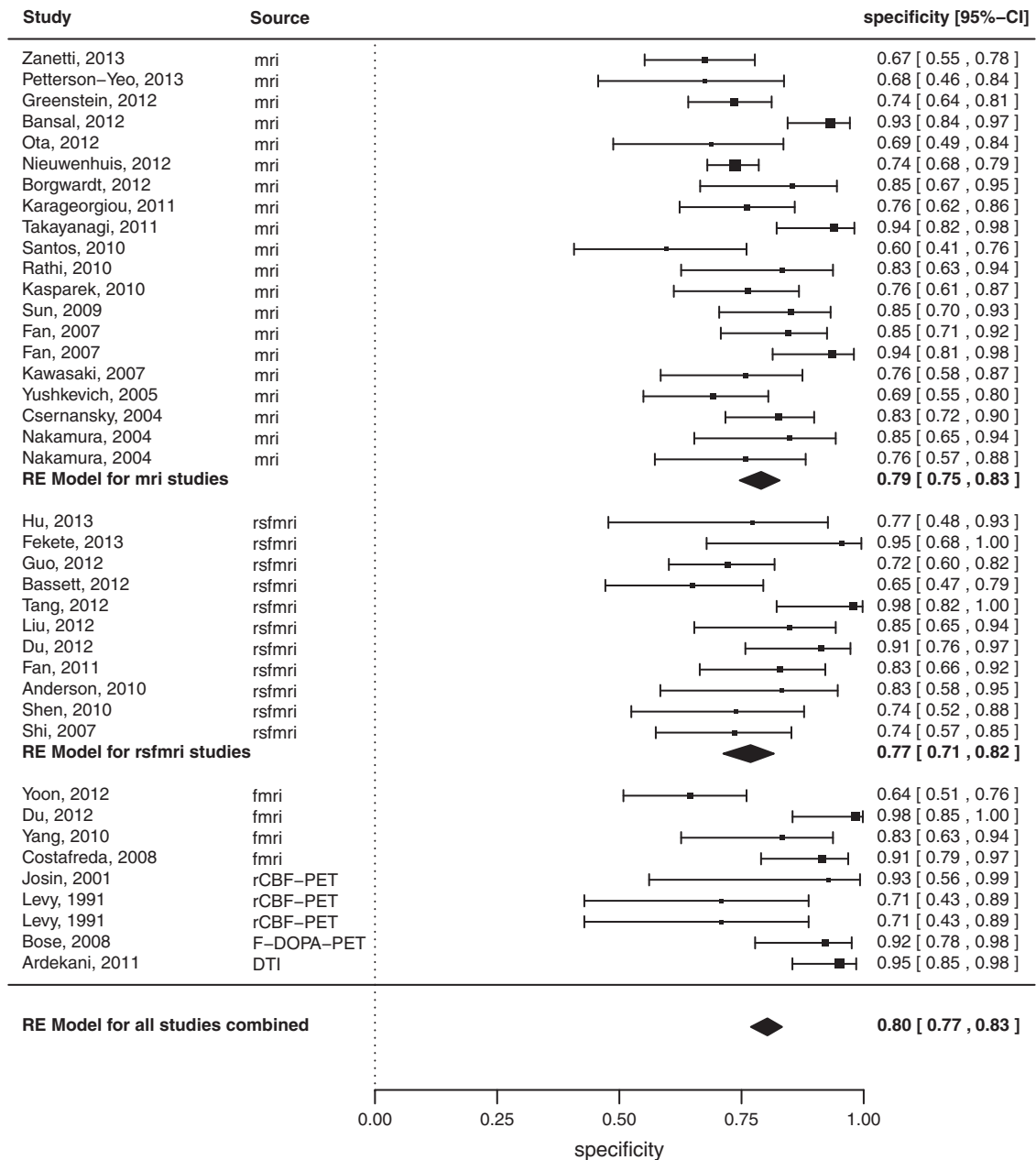
| Study | Source | | specificity [95%–CI] |
|---|---|---|---|
| Zanetti, 2013 | mri | | 0.67 [ 0.55 , 0.78 ] |
| Petterson–Yeo, 2013 | mri | | 0.68 [ 0.46 , 0.84 ] |
| Greenstein, 2012 | mri | | 0.74 [ 0.64 , 0.81 ] |
| Bansal, 2012 | mri | | 0.93 [ 0.84 , 0.97 ] |
| Ota, 2012 | mri | | 0.69 [ 0.49 , 0.84 ] |
| Nieuwenhuis, 2012 | mri | | 0.74 [ 0.68 , 0.79 ] |
| Borgwardt, 2012 | mri | | 0.85 [ 0.67 , 0.95 ] |
| Karageorgiou, 2011 | mri | | 0.76 [ 0.62 , 0.86 ] |
| Takayanagi, 2011 | mri | | 0.94 [ 0.82 , 0.98 ] |
| Santos, 2010 | mri | | 0.60 [ 0.41 , 0.76 ] |
| Rathi, 2010 | mri | | 0.83 [ 0.63 , 0.94 ] |
| Kasparek, 2010 | mri | | 0.76 [ 0.61 , 0.87 ] |
| Sun, 2009 | mri | | 0.85 [ 0.70 , 0.93 ] |
| Fan, 2007 | mri | | 0.85 [ 0.71 , 0.92 ] |
| Fan, 2007 | mri | | 0.94 [ 0.81 , 0.98 ] |
| Kawasaki, 2007 | mri | | 0.76 [ 0.58 , 0.87 ] |
| Yushkevich, 2005 | mri | | 0.69 [ 0.55 , 0.80 ] |
| Csernansky, 2004 | mri | | 0.83 [ 0.72 , 0.90 ] |
| Nakamura, 2004 | mri | | 0.85 [ 0.65 , 0.94 ] |
| Nakamura, 2004 | mri | | 0.76 [ 0.57 , 0.88 ] |
| **RE Model for mri studies** | | | **0.79 [ 0.75 , 0.83 ]** |
| | | | |
| Hu, 2013 | rsfmri | | 0.77 [ 0.48 , 0.93 ] |
| Fekete, 2013 | rsfmri | | 0.95 [ 0.68 , 1.00 ] |
| Guo, 2012 | rsfmri | | 0.72 [ 0.60 , 0.82 ] |
| Bassett, 2012 | rsfmri | | 0.65 [ 0.47 , 0.79 ] |
| Tang, 2012 | rsfmri | | 0.98 [ 0.82 , 1.00 ] |
| Liu, 2012 | rsfmri | | 0.85 [ 0.65 , 0.94 ] |
| Du, 2012 | rsfmri | | 0.91 [ 0.76 , 0.97 ] |
| Fan, 2011 | rsfmri | | 0.83 [ 0.66 , 0.92 ] |
| Anderson, 2010 | rsfmri | | 0.83 [ 0.58 , 0.95 ] |
| Shen, 2010 | rsfmri | | 0.74 [ 0.52 , 0.88 ] |
| Shi, 2007 | rsfmri | | 0.74 [ 0.57 , 0.85 ] |
| **RE Model for rsfmri studies** | | | **0.77 [ 0.71 , 0.82 ]** |
| | | | |
| Yoon, 2012 | fmri | | 0.64 [ 0.51 , 0.76 ] |
| Du, 2012 | fmri | | 0.98 [ 0.85 , 1.00 ] |
| Yang, 2010 | fmri | | 0.83 [ 0.63 , 0.94 ] |
| Costafreda, 2008 | fmri | | 0.91 [ 0.79 , 0.97 ] |
| Josin, 2001 | rCBF–PET | | 0.93 [ 0.56 , 0.99 ] |
| Levy, 1991 | rCBF–PET | | 0.71 [ 0.43 , 0.89 ] |
| Levy, 1991 | rCBF–PET | | 0.71 [ 0.43 , 0.89 ] |
| Bose, 2008 | F–DOPA–PET | | 0.92 [ 0.78 , 0.98 ] |
| Ardekani, 2011 | DTI | | 0.95 [ 0.85 , 0.98 ] |
| | | | |
| **RE Model for all studies combined** | | | **0.80 [ 0.77 , 0.83 ]** |

specificity scale: 0.00  0.25  0.50  0.75  1.00

**Figure 2** Forest plot of specificities for studies using MRI, fMRI, rsfMRI, rCBF-PET, F-DOPA-PET, and DTI to diagnose schizophrenia. Summary estimates for specificity are computed using the approach described by Reitsma et al (2005).

(Tost et al, 2010) and long-term (Navari and Dazzan, 2009) antipsychotic treatment. Thus, pronounced brain changes and higher sensitivity of neuroimaging-based diagnostic models in older patients might additionally result from long-standing antipsychotic treatment (Fusar-Poli et al, 2013; Ho et al, 2011; Smieskova et al, 2009). The investigation of antipsychotic treatment as a moderator in the present analysis indicated a potential effect of the current antipsychotic dose. However, while older age was associated with higher sensitivity, higher chlorpromazine equivalents were associated with higher specificity. To further disentangle possible effects of antipsychotic medication on diagnostic classification measures from the impact of the disease process itself, future meta-analyses have to

cover a critical mass of patient samples having well-documented prospective medication data (Ho et al, 2011).

## Effect of Psychotic Symptoms

Another interesting finding of the present analysis is the association between predominant positive symptoms and higher specificity of the neuroimaging-based diagnostic models. It has been reported that brain changes associated with schizophrenia are related to the extent of psychopathology as measured by psychotic symptom scales (Modinos et al, 2013; Palaniyappan et al, 2012). Similarly, there seem to be differences in brain alterations in patients with predominant positive vs predominant negative

symptoms (Koutsouleris *et al*, 2008; Nenadic *et al*, 2010). This might seem counterintuitive as previous studies indicate larger brain structural abnormalities in patients with pronounced negative symptom symptoms (Koutsouleris *et al*, 2008). However, it might be the case that the pattern of gray matter alterations in patients with mainly positive symptoms—even if it is subtle—is more distinctive as compared with patients with negative symptoms and thus facilitates higher classification performances. It may be hypothesized that patients with predominant positive symptoms also received higher dosages of antipsychotic medication, which in turn may impact on the brain as discussed above. Therefore, the finding that positive symptoms are associated with specificity might be confounded by previous treatment. Another potential interpretation of this association may relate to the current, purely symptom-based diagnostic system, which forms the ground truth for fully supervised neuroimaging-based disease classification. In this regard, greater homogeneity between clinical raters can be expected when they diagnose schizophrenia in patients with pronounced positive symptoms as compared to patients with negative symptoms, who are difficult to differentiate from patients with major or psychotic depression. Thus, predominant negative symptoms might be associated with higher neurobiological variability compared with the phenotype



**Figure 3** SROC curve of the Reitsma model with the summary sensitivity and false positive rate indicated in black as well as color-coded the sensitivity and false positive rate of the invidivual studies of different imaging modalities.

of acute psychosis, creating an area of diagnostic ambiguity not only for clinical raters but also for any downstream supervised classification methods relying on these raters. In fact, this would create an upper bound on the sensitivity and specificity that could be achieved by means of supervised neuroimaging-based predictive models.

## Effect of Neuroimaging Modality

Our results point to a significantly higher sensitivity associated with rsfMRI data as compared with structural MRI data whereas both neuroimaging modalities showed a similar specificity. This suggests that more homogeneous functional resting-state patterns in schizophrenia lead to a tighter clustering of patients in the functional compared with the structural feature space, and hence to an increased capacity of rsfMRI-based classification algorithms to detect the disease condition. This in turn suggests that disease heterogeneity is greater in the neuroanatomical domain as show in a recent study (Zhang *et al*, 2014). Future studies may involve both structural and functional MRI data to generate diagnostic classifiers with superior sensitivity and specificity.

## Differences between Multivariate Methods

We observed a substantial methodological heterogeneity concerning the multivariate algorithms used to build the predictive models. The most frequent approaches were discriminant analysis and support-vector machines, which were used by 26 out of 36 studies (72%). It is important to note that support-vector machines typically show higher classification performance when nonlinear relationships are present in the data. Also—unlike linear discriminant analysis—support-vector machines increase generalizability by emphasizing samples that are located close to the decision boundary in the feature space (Hastie *et al*, 2009). As both approaches showed almost identical sensitivity and specificity in our analysis, these differences seem not to have affected the classification performance. Three studies (Bose *et al*, 2008; Josin and Liddle, 2001; Rathi *et al*, 2010) applied an artificial neural network model to structural MRI and PET data with slightly higher sensitivity (86–100%) and a slightly higher specificity (85–100%). Two studies (Anderson *et al*, 2010; Greenstein *et al*, 2012) applied a random-forest approach to fMRI and MRI data. These studies report a slightly lower sensitivity (64 and 73%) and slightly lower specificity (83 and 74%) compared with other studies. However, it is noteworthy that the comparison of different classification methods in the context of the present meta-analysis might be confounded by the characteristics of
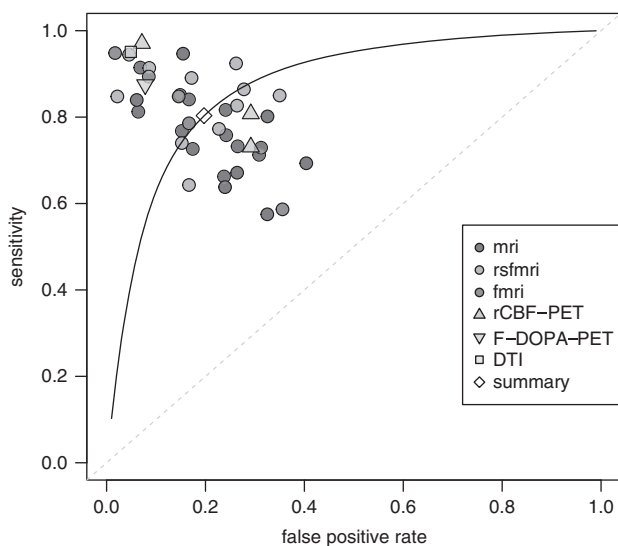
**Table 1** Results from Bivariate Meta-analyses Applying the Approach by Reitsma *et al* (2005)

| Data | N of controls | N of patients | Sensitivity | Specificity | Positive LR | Negative LR | Diagnostic OR |
|---|---|---|---|---|---|---|---|
| MRI | 1030 | 988 | 76.42 (71.87–80.44) | 79.01 (74.63–82.81) | 3.67 (2.88–4.61) | 0.3 (0.24–0.37) | 12.5 (7.88–18.9) |
| rsfMRI | 275 | 315 | 84.46 (79.86–88.17) | 76.85 (71.28–81.63) | 3.67 (2.92–4.61) | 0.204 (0.153–0.265) | 18.5 (11.8–27.7) |
| All combined | 1602 | 1637 | 80.34 (76.71–83.52) | 80.3 (76.92–83.29) | 4.1 (3.37–4.93) | 0.246 (0.2–0.298) | 17 (11.5–24.2) |

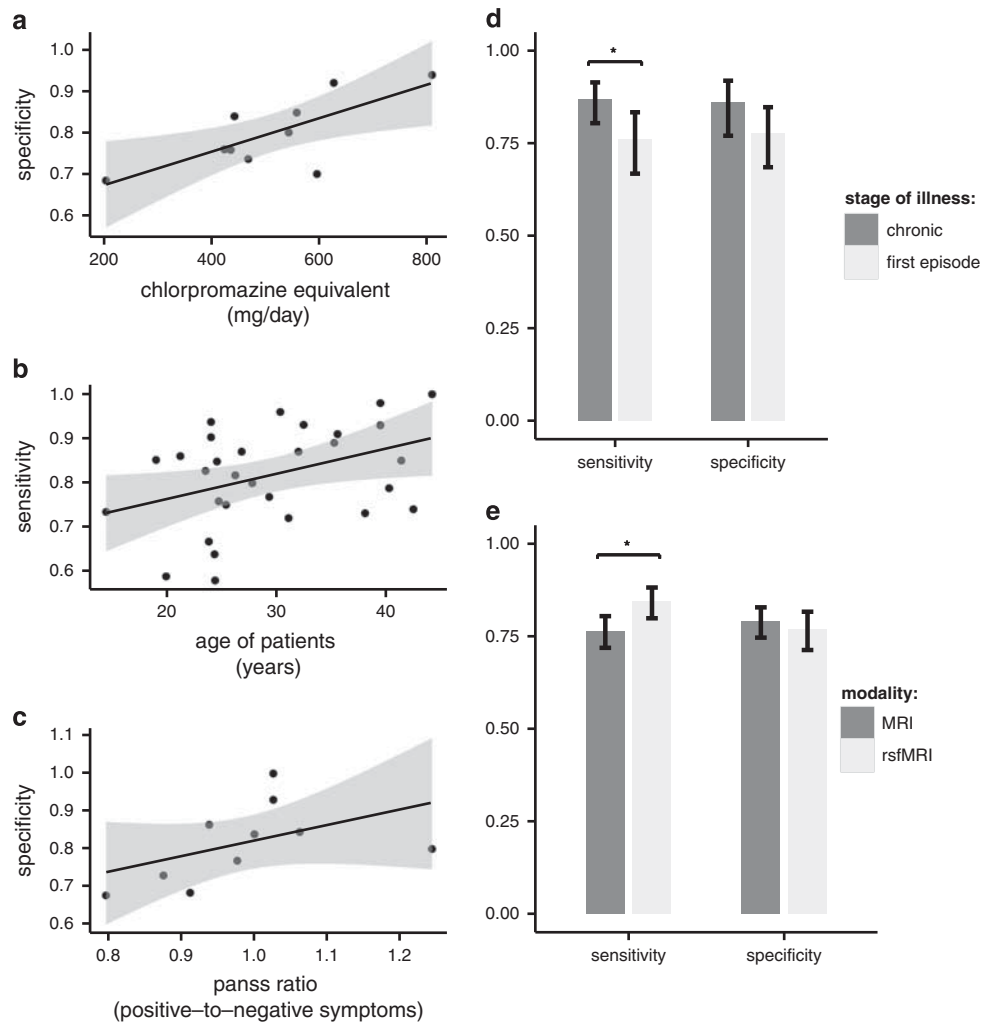Positive LR, negative LR, and DOR are estimated via MCMC (Zwinderman and Bossuyt, 2008).

**Figure 4** Results from the moderator analysis: linear regression models with (a) chlorpromazin equivalent predicting specificity, (b) age of patients predicting sensitivity, (c) PANSS ratio predicting specificity and differences in sensitivity and specificity between (d) stages of illness and (e) imaging modalities.

the investigated samples such as age, medication, symptoms, and disease stage. To the best of our knowledge, a systematic investigation of different classification algorithms for the MRI-based diagnosis of schizophrenia in large representative patient populations is missing.

### Limitations of the Presented Study

It is of note that $\sim 20\%$ of patients were misclassified as HC by the applied multivariate models. This misclassification rate may be due to either (1) the existence of a different pattern of brain abnormalities in this subgroup compared with the majority of patients, or (2) to the absence of a homogenous discriminative pattern in this patient subgroup compared to the HC group, or (3) to the rater-based 'noise' in the diagnostic labels provided to supervised classification algorithms. The aggregated data analysis performed in our study does not allow us to clarify these alternative possibilities. Hence, future studies employing semi- and unsupervised machine learning methods in well-controlled representative study populations are needed to potentially elucidate the neurobiological heterogeneity

of the disorder and in turn use this information to generate high-performing neuroimaging-based classifiers of schizophrenia.

In this context, it needs to be noted that most of the published studies on neuroimaging-based diagnostic models largely focus on methodological details of the applied machine learning algorithms. This results from the fact that multivariate prediction of psychiatric diagnosis is a young research topic. Thus, most studies aim at 'proof of concept' approaches, showing that multivariate models are principally able to infer distinctive brain patterns at the single-subject level. Another reason might be the availability of numerous competing algorithms. Most studies so far have tried to compare new techniques to previous ones while paying little attention to the systematic investigation of methodological factors within the same sample.

On the other hand, most studies provide only limited information regarding the investigated patients samples and their clinical characteristics. As pointed out by Deville et al (Devillé et al, 2002), a detailed description of the patients' disease status, symptoms, length and course of

illness, current medication, or comorbidities is crucial for evaluating the potential of such models to enter clinical practice in the future. The results of our meta-analysis fully agree with this requirement as they showed that clinical factors such as age or symptoms affect sensitivity and specificity, while methodological factors did not. As such some patient samples might be more suitable for the application of neuroimaging-based predictive models than others. This also has implications for the interpretation of neuroimaging-based predictive models. There are multiple confounding factors that are illness-related, but not causative, that might result in neurobiological differentiation. Thus, to move from a theoretical field of research toward a clinical application of these diagnostic methods, future studies should provide detailed clinical and sociodemographic information about the investigated patient and HC samples. This clinical information is the 'conditio sine qua non' for evaluating the applicability of multivariate methods to various patient samples, subsamples or disease states.

It must be noted that the studies included in the present analysis identified schizophrenia-distinctive brain patterns as compared to healthy volunteers. To date, only few studies have investigated patterns of brain abnormalities that differentiate between different psychiatric disorders. For the differentiation between schizophrenic and bipolar patients diagnostic accuracies of 92% for schizophrenia and 79% for bipolar disorder based on fMRI (Costafreda et al, 2011) and an overall classification accuracy of 88% (Schnack et al, 2014) or 100% (Bansal et al, 2012) based on sMRI have been reported. This research direction is critical as there is considerable doubt whether the current nosological constructs are subserved by distinct neurobiological signatures, or alternatively whether there exists a significant pathophysiological overlap between disease entities. A promising strategy to address this issue might be the delineation of more homogenous patient subgroups within and across disease boundaries (Insel et al, 2010) by means of unsupervised and semisupervised analysis methods (Filipovych et al, 2011, 2012). Also, future studies need to address the question of how well neuroimaging-based biomarkers generalize, eg, across different sites. In the studies included in the present analysis most of the data have been acquired on one site using the same scanners and scanning sequences. However a recent study indicates that diagnostic models are not site specific but that similar sensitivity and specificity can be achieved for data acquired from different sites (Nieuwenhuis et al, 2012). Although the present analysis indicates a discriminative pattern of brain alterations associated with schizophrenia, our results underline the importance of an exhaustive assessment of clinical characteristics during the investigation of such biomarkers.

## FUNDING AND DISCLOSURE

## ACKNOWLEDGEMENTS

## REFERENCES

Anderson A, Dinov ID, Sherin JE, Quintana J, Yuille AL, Cohen MS (2010). Classification of spatially unaligned fMRI scans. NeuroImage 49: 2509–2519.

Bansal R, Staib LH, Laine AF, Hao X, Xu D, Liu J et al (2012). Anatomical brain images alone can accurately diagnose chronic neuropsychiatric illnesses. PloS One 7: e50698.

Borgwardt S, Fusar-Poli P (2012). Third-generation neuroimaging in early schizophrenia: translating research evidence into clinical utility. Br J Psychiatry 200: 270–272.

Borgwardt S, Radua J, Mechelli A, Fusar-Poli P (2012). Why are psychiatric imaging methods clinically unreliable? Conclusions and practical guidelines for authors, editors and reviewers. Behav Brain Funct BBF 8: 46.

Bose SK, Turkheimer FE, Howes OD, Mehta MA, Cunliffe R, Stokes PR et al (2008). Classification of schizophrenic patients and healthy controls using [18 F] fluorodopa PET imaging. Schizophr Res 106: 148–155.

Castellani U, Rossato E, Murino V, Bellani M, Rambaldelli G, Perlini C et al (2012). Classification of schizophrenia using feature-based morphometry. J Neural Transm Vienna Austria 1996 119: 395–404.

Chan RCK, Di X, McAlonan GM, Gong Q-Y (2009). Brain anatomical abnormalities in high-risk individuals, first-episode, and chronic schizophrenia: an activation likelihood estimation meta-analysis of illness progression. Schizophr Bull 37: 177–188.

Costafreda SG, Fu CHY, Picchioni M, Toulopoulou T, McDonald C, Kravariti E et al (2011). Pattern of neural responses to verbal fluency shows diagnostic specificity for schizophrenia and bipolar disorder. BMC Psychiatry 11: 18.

Davatzikos C (2004). Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. NeuroImage 23: 17–20.

Davatzikos C, Shen D, Gur RC, Wu X, Liu D, Fan Y et al (2005). Whole-brain morphometric study of schizophrenia revealing a spatially complex set of focal abnormalities. Arch Gen Psychiatry 62: 1218–1227.

Deeks JJ, Macaskill P, Irwig L (2005). The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 58: 882–893.

Devillé WL, Buntinx F, Bouter LM, Montori VM, de Vet HC, van der Windt DA et al (2002). Conducting systematic reviews of diagnostic studies: didactic guidelines. BMC Med Res Methodol 2: 9.

Doebler P (2012): Meta-Analysis of Diagnostic Accuracy with mada. Available at < http://cran.gis-lab.info/web/packages/mada/vignettes/mada.pdf>. Last accessed date 1 August 2014.

Fan Y, Shen D, Gur RC, Gur RE, Davatzikos C (2007). COMPARE: classification of morphological patterns using adaptive regional elements. IEEE Trans Med Imaging 26: 93–105.

Fekete T, Wilf M, Rubin D, Edelman S, Malach R, Mujica-Parodi LR (2013). Combining classification with fMRI-derived complex

network measures for potential neurodiagnostics. *PloS One* **8**: e62867.

Filipovych R, Davatzikos CAlzheimer's Disease Neuroimaging Initiative (2011). Semi-supervised pattern classification of medical images: application to mild cognitive impairment (MCI). *NeuroImage* **55**: 1109–1119.

Filipovych R, Resnick SM, Davatzikos C (2012). JointMMCC: joint maximum-margin classification and clustering of imaging data. *IEEE Trans Med Imaging* **31**: 1124–1140.

Fornito A, Yücel M, Patti J, Wood SJ, Pantelis C (2009). Mapping grey matter reductions in schizophrenia: an anatomical likelihood estimation analysis of voxel-based morphometry studies. *Schizophr Res* **108**: 104–113.

Fu CHY, Costafreda SG (2013). Neuroimaging-based biomarkers in psychiatry: clinical opportunities of a paradigm shift. *Can J Psychiatry Rev Can Psychiatr* **58**: 499–508.

Fusar-Poli P, Smieskova R, Kempton MJ, Ho BC, Andreasen NC, Borgwardt S (2013). Progressive brain changes in schizophrenia related to antipsychotic treatment? A meta-analysis of longitudinal MRI studies. *Neurosci Biobehav Rev* **37**: 1680–1691.

Gatsonis C, Paliwal P (2006). Meta-analysis of diagnostic and screening test accuracy evaluations: methodologic primer. *AJR Am J Roentgenol* **187**: 271–281.

Greenstein D, Malley JD, Weisinger B, Clasen L, Gogtay N (2012). Using multivariate machine learning methods and structural MRI to classify childhood onset schizophrenia and healthy controls. *Front Psychiatry Front Res Found* **3**: 53.

Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JAC (2007). A unification of models for meta-analysis of diagnostic accuracy studies. *Biostat Oxf Engl* **8**: 239–251.

Hastie T, Tibshirani R, Friedman J (2009). *The elements of statistical learning*. Springer, **2**.

Ho B-C, Andreasen NC, Ziebell S, Pierson R, Magnotta V (2011). Long-term antipsychotic treatment and brain volumes: a longitudinal study of first-episode schizophrenia. *Arch Gen Psychiatry* **68**: 128–137.

Honea R, Crow TJ, Passingham D, Mackay CE (2005). Regional deficits in brain volume in schizophrenia: a meta-analysis of voxel-based morphometry studies. *Am J Psychiatry* **162**: 2233–2245.

Howes OD, Kambeitz J, Kim E, Stahl D, Slifstein M, Abi-Dargham A *et al* (2012). The nature of dopamine dysfunction in schizophrenia and what this means for treatment: meta-analysis of imaging studies. *Arch Gen Psychiatry* **69**: 776–786.

Hu X, Zhu D, Lv P, Li K, Han J, Wang L *et al* (2013). Fine-granularity functional interaction signatures for characterization of brain conditions. *Neuroinformatics* **11**: 301–317.

Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K *et al* (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry* **167**: 748–751.

Ioannidis J, Lau J (2001). Evolution of treatment effects over time: empirical insight from recursive cumulative metaanalyses. *Proc Natl Acad Sci USA* **98**: 831–836.

Josin GM, Liddle PF (2001). Neural network analysis of the pattern of functional connectivity between cerebral areas in schizophrenia. *Biol Cybern* **84**: 117–122.

Kapur S, Phillips AG, Insel TR (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatry* **17**: 1174–1179.

Kawasaki Y, Suzuki M, Kherif F, Takahashi T, Zhou S-Y, Nakamura K *et al* (2007). Multivariate voxel-based morphometry successfully differentiates schizophrenia patients from healthy controls. *NeuroImage* **34**: 235–242.

Koutsouleris N, Gaser C, Jäger M, Bottlender R, Frodl T, Holzinger S *et al* (2008). Structural correlates of psychopathological symptom dimensions in schizophrenia: a voxel-based morphometric study. *NeuroImage* **39**: 1600–1612.

Liu M, Zeng L-L, Shen H, Liu Z, Hu D (2012). Potential risk for healthy siblings to develop schizophrenia: evidence from pattern classification with whole-brain connectivity. *Neuroreport* **23**: 265–269.

McGrath J, Saha S, Chant D, Welham J (2008). Schizophrenia: a concise overview of incidence, prevalence, and mortality. *Epidemiol Rev* **30**: 67–76.

McIntosh AR, Lobaugh NJ (2004). Partial least squares analysis of neuroimaging data: applications and advances. *NeuroImage* **23**(Suppl 1): S250–S263.

Minzenberg MJ, Laird AR, Thelen S, Carter CS, Glahn DC (2009). Meta-analysis of 41 functional neuroimaging studies of executive function in schizophrenia. *Arch Gen Psychiatry* **66**: 811–822.

Modinos G, Costafreda SG, van Tol M-J, McGuire PK, Aleman A, Allen P (2013). Neuroanatomy of auditory verbal hallucinations in schizophrenia: a quantitative meta-analysis of voxel-based morphometry studies. *Cortex* **49**: 1046–1055.

Moher D, Liberati A, Tetzlaff J, Altman DG (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* **339**: b2535.

Navari S, Dazzan P (2009). Do antipsychotic drugs affect brain structure? A systematic and critical review of MRI findings. *Psychol Med* **39**: 1763–1777.

Nenadic I, Sauer H, Gaser C (2010). Distinct pattern of brain structural deficits in subsyndromes of schizophrenia delineated by psychopathology. *NeuroImage* **49**: 1153–1160.

Nieuwenhuis M, Haren NEM, van, Hulshoff Pol HE, Cahn W, Kahn RS, Schnack HG (2012). Classification of schizophrenia patients and healthy controls from structural MRI scans in two large independent samples. *NeuroImage* **61**: 606–612.

Palaniyappan L, Balain V, Radua J, Liddle PF (2012). Structural correlates of auditory hallucinations in schizophrenia: a meta-analysis. *Schizophr Res* **137**: 169–173.

Pers TH, Albrechtsen A, Holst C, Sørensen TIA, Gerds TA (2009). The validation and assessment of machine learning: a game of prediction from high-dimensional data. *PLoS One* **4**: e6287.

Rathi Y, Malcolm J, Michailovich O, Goldstein J, Seidman L, McCarley RW *et al* (2010). Biomarkers for identifying first-episode schizophrenia patients using diffusion weighted imaging. *Med Image Comput Comput-Assist Interv* **13**: 657–665.

R Core Team (2013). R Foundation for Statistical Computing: Vienna, Austria, R: A Language and Environment for Statistical Computing. Available at <http://www.R-project.org/>.

Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* **58**: 982–990.

Rutter CM, Gatsonis CA (2001). A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* **20**: 2865–2884.

Schnack HG, Nieuwenhuis M, Haren NEM, van, Abramovic L, Scheewe TW, Brouwer RM *et al* (2014). Can structural MRI aid in clinical classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects. *NeuroImage* **84**: 299–306.

Shen H, Wang L, Liu Y, Hu D (2010). Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI. *NeuroImage* **49**: 3110–3121.

Smieskova R, Fusar-Poli P, Allen P, Bendfeldt K, Stieglitz RD, Drewe J *et al* (2009). The effects of antipsychotics on the brain: what have we learnt from structural imaging of schizophrenia? A systematic review. *Curr Pharm Des* **15**: 2535–2549.

Tost H, Braus DF, Hakimi S, Ruf M, Vollmert C, Hohn F *et al* (2010). Acute D2 receptor blockade induces rapid, reversible remodeling in human cortical-striatal circuits. *Nat Neurosci* **13**: 920–922.

Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB *et al* (2011). QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* **155**: 529–536.

WHO (2004). Global Burden of Disease—Update. Available at <http://www.who.int/healthinfo/global_burden_disease/en>.

Zarogianni E, Moorhead TWJ, Lawrie SM (2013). Towards the identification of imaging biomarkers in schizophrenia, using multivariate pattern classification at a single-subject level. *NeuroImage Clin* **3**: 279–289.

Zhang T, Koutsouleris N, Meisenzahl E, Davatzikos C (2014). Heterogeneity of structural brain changes in subtypes of schizophrenia revealed using magnetic resonance imaging pattern analysis. *Schizophr Bull* **41**: 74–84.

Zwinderman A, Bossuyt P (2008). We should not pool diagnostic likelihood ratios in systematic reviews. *Stat Med* **27**: 687–697.

Supplementary Information accompanies the paper on the Neuropsychopharmacology website (http://www.nature.com/npp)