Detecting neuroimaging biomarkers for schizophrenia

Kambeitz, Joseph; Kambeitz-Ilankovic, Lana; Leucht, Stefan; Wood, Stephen; Davatzikos, Christos; Malchow, Berend; Falkai, Peter; Koutsouleris, Nikolaos

DOI:
10.1038/npp.2015.22

Abstract

Introduction:

The introduction of multivariate statistical methods in the analysis of neuroimaging data allowed the generation of highly predictive models. These allow the differentiation of patients from healthy controls solely based on neuroimaging data. However there exist substantial heterogeneity in the reported accuracies making it difficult to evaluate the overall potential of these studies to inform clinical diagnosis.

Methodology:

We conducted a comprehensive literature search to identify all studies that used multivariate statistical methods to differentiate patients with schizophrenia from healthy controls based solely on neuroimaging data. A bivariate random-effects meta-analytic model was implemented to investigate diagnostic accuracy across studies. The robustness of the results as well as the effect of potentially confounding continous variables was investigated by moderator analysis.

Results:

Meta-analysis of the complete sample (n=36 studies, n=1525 patients, n=1536 healthy controls) showed a sensitivity of 80.7% (95%-CI: 77.0 to 83.9%) and a specificity of 80.2% (95%-CI: 83.3 to 76.7%). Moderator analysis showed significant effects of age of patients (p=0.021), imaging modality (p=0.019) and stage of disease (p=0.003) on sensitivity as well as of positive-to-negative symptom ratio (p=0.028) and chlorpromazine equivalent (p=0.016) on specificity.

Discussion:

Our analysis indicate an overall sensitivity and specificity of around 80 % of neuroimaging-based predictive models for differentiating schizophrenic patients from healthy controls. However the diagnostic accuracy is affected by several potentially confounding factors such as the age of patients, their disease stage, previous medication as well as clinical symptoms.

Key words: psychosis, classification, prediction, sensitivity, specificity, diagnosis

Introduction

Schizophrenia shows prevalence rates of 0.5-1% in the general population making it one of the leading factors of global disease burden (WHO, 2004). Diagnosis of schizophrenia is based on the psychiatrist‘s evaluation of patients’ reported symptoms and behaviour. Substantial efforts have been invested in identifying biomarkers that can potentially guide the diagnostic process and allow more accurate diagnosis than the current state-of-the-art. As an example, multiple functional[1,2] (Kambeitz et al., in press) and structural brain changes[3–5] have been associated with schizophrenia. Even though these results indicate significant differences in e.g. brain structure between healthy controls and patients at the *group level* - there is a substantial overlap between groups. This makes brain changes unsuitable to guide diagnosis at the *individual level*. Therefore alterations in brain structure and function have not been successfully integrated into the diagnostic process as disease markers for the individual subject[6–8]. Traditional univariate or mass-univariate statistical approaches neglect the heavily interconnected nature of functional and structural brain data[9]. An increasing number of studies have applied novel multivariate statistical approaches to the analysis of brain alterations in patients with schizophrenia (e.g.[10,11]). The results indicate that a constellation of subtle structural or functional changes can be highly distinctive of schizophrenia-related brain change, even though each individual component might not.

Most importantly multivariate statistical methods can provide predictive models that allow for the diagnostic classification of unseen individuals. This might open up the possibility of neuroimaging contributing to the clinical diagnostic process. For instance, support vector machines[11], partial least squares analysis[10,12], random forests[13,14] and artificial neural networks[15–17] have been shown to differentiate patients from healthy controls with diagnostic accuracies of up to 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 diagnostic accuracy of the reported predictive models differ largely, making it difficult to evaluate the overall potential of these studies to inform clinical diagnosis. Little is known about which factors determine the success of neuroimaging-based predictive modeling. Some studies have compared two or more algorithms[16–18]. However, a systematic investigation of different imaging modalities or multivariate methods is still missing. No comparative reports exist on the relationship between clinical variables of the tested samples and diagnostic accuracies of neuroimaging-based predictive models. Age, gender, psychiatric symptoms or current medication represent potentially confounding variables, which might affect the diagnostic success of such models.

Thus, to shed light on the potential application of multivariate models for disease classification, we conducted a meta-analysis of the performance of these models when applied to neuroimaging data of patients with schizophrenia and healthy controls, as measured by their overall diagnostic accuracy. Within this framework, we evaluated the potentially moderating impact of different clinical variables on the observed diagnostic accuracies.

Methods

Search and selection strategy

The entire PubMed electronic database was searched from 1st January 1950 up to 31st May 2013. Initially, studies were screened using a comprehensive search term [("support vector" OR "SVM" OR "classification" OR "categorization") AND ("MRI" OR "fMRI" OR "magnetic resonance" OR "imaging" OR "grey 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 healthy controls. 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 was available that allowed the calculation of the following parameters: true positives (TP), true negatives (TN), false positives (FP), false negatives (FN). In case insufficient data was reported, authors were contacted via email to provide additional information regarding their published reports. In multivariate classification it is of high importance to apply some form of crossvalidation while estimating model parameters to avoid overfitting, which is associated with low generalizability ("avoid double dipping"[19]). Thus, only studies that applied crossvalidation (e.g. leave-one-out, n-fold, out-of-bags) were included in the analysis. In some cases there were multiple published studies 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[20] (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 healthy controls as measured by sensitivity ($sensitivity = \frac{number\ of\ true\ positives}{number\ of\ true\ positives + number\ of\ false\ negatives}$), specificity ($specificity = \frac{number\ of\ true\ negatives}{number\ of\ true\ negatives + number\ of\ false\ positives}$), true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The following additional information was extracted from all studies: names of the authors, year of publication, population characteristics of the healthy control and patient groups (group size, age, gender, antipsychotic use, diagnosis, 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, resolution), characteristics of the employed preprocessing methodology, characteristics of the classification procedure (e.g. linear discriminant analysis,

support vector machine) and characteristics of the crossvalidation 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)[21].

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[22]. Instead a bivariate approach[23] and an approach based on a hierarchical summary ROC model (HSROC[24]) have been suggested to estimate diagnostic accuracy across studies. However, in most situations both approaches lead to identical results[25]. In the present analysis we implemented the approach by Reitsma et al.[23]. 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 due to differences in study populations, sampling errors, and differences in implicit thresholds applied to the data to separate patients from healthy controls. 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. Results from the meta-analysis are presented in forest plots separately for sensitivity and specificity. Summary estimates for sensitivity and specificity are provided separately for MRI, for fMRI, for rs-fMRI 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[26]. The robustness of the results as well as the effect of potentially confounding variables (e.g. age, gender ratio, year of publication) was investigated by adding moderator variables to the bivariate regression model. Publication bias was assessed via computation of log diagnostic odds ratios (logDOR) for all studies as suggested by Deeks et al.[27]. LogDOR are plotted against the effective sample size in order to create funnel plots[27]. Symmetry of funnel plots was investigated via visual inspection. All computations were performed using the R statistical programming language version 2.10.13[28] with the package mada[29].

Results

The initial literature search identified 397 studies of interest. After screening all studies and applying the inclusion criteria, 360 studies were excluded. Fan et al.[30] and Davatzikos et al.[11] used overlapping samples. Only Fan et al.[30] 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[11]. Between Liu et al.[31] and Shen et al.[32] 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. The final sample consisted of n=36 studies with of a total of n=1525 patients with schizophrenia and n=1536 healthy controls. Among the included studies were n=19 studies using structural MRI, n=7 studies using rsfMRI, n=6 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).

For all studies combined there was a sensitivity of 80.7% (95%-CI: 77.0 to 83.9%, see Figure 1) and a specificity of 80.2% (95%-CI: 83.3 to 76.7%, see Figure 2). A summary ROC-curve of the included studies as well as the estimated summary is presented in Figure 3. Visual inspection of a funnel plot 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=0.866) or specificity (p=0.812).

There was no significant effect of gender ratio in patients or controls, illness duration, PANSS positive scores, PANSS negative scores or analysis method (SVM/LDA) on either sensitivity of specificity (all $p > 0.1$). There was a significant effect of age of patients on sensitivity (p=0.021) indicating better diagnostic accuracy in older patients (see Figure 4). There was no evidence for an effect of patient's age on specificity (p=0.207) and no effect of age of the healthy controls on sensitivity (p=0.114) of specificity (p=0.494). There was a significant effect of positive-to-negative symptom ratio on specificity (p=0.030), 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=0.805). Comparing studies that investigate first-episode patients with patients in a chronic stage of schizophrenia, there was a significantly higher sensitivity in patients in a chronic stage (p=0.003, see Figure 4) but no effect on specificity (p=0.235). There was a significant effect of the chlorpromazin equivalent of the investigated samples on specificity (p=0.016) indicating higher specificity in subjects with higher dosis of medication (see Figure 4). Chlorpromazine equivalent did not effect sensitivity (p=0.06). When the structural MRI studies were analyzed separately the meta-analysis showed a sensitivity of 77.3% (95%-CI: 72.7 to 81.3%) and a specificity of 78.7% (95%-CI: 82.3 to 74.5%). For the fMRI studies it showed a sensitivity of 81.4% (95%-CI: 67.3 to 90.2%) and a specificity of 82.4% (95%-CI: 90.5 to 69.6%). For only the rs-fMRI studies there was a sensitivity of 86.9% (95%-CI: 81.4 to 91.0%) and a specificity of 80.3% (95%-CI: 88.1 to 69.2%). "Data source" was added as a moderating variable to the bivariate meta-analysis model to investigate significant differences between different data sources (MRI, fMRI, rs-fMRI). There was a significant difference (p=0.019) between the sensitivity of rs-fMRI and MRI studies, indicating higher sensitivity in rs-fMRI studies (see Figure 4). There was no significant difference in sensitivity and specificity between any other source of data (see figure). In order 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 no significant difference between DA and SVM studies regarding sensitivity (p=0.766) and specificity (p=0.801).

Comment [nk1]: I would have expected patients with negative symptoms tob e easier classifier .

Comment [nk2]: I would have expected patients with negative symptoms tob e easier classifier .

Comment [SJW3]: But schizophrenia is not diagnosed on the basis of negative symptoms! So one would expect the positive symptoms to be better classifiers both when looking at symptoms and when looking at the neurobiological basis of the diagnosis

Discussion

We present a meta-analysis of a total of n=36 studies including n=1525 patients with schizophrenia and n=1536 healthy controls are presented. Our results indicate an overall sensitivity and overall specificity of around 80% of neuroimaging-based predictive models for differentiating schizophrenic patients from healthy controls. Similar results were obtained when analysis was restricted to individual imaging modalities (structural MRI, fMRI, or rs-fMRI). Also 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 of the investigated subjects was significantly associated with higher sensitivity. While illness duration did not have a significant impact on diagnostic accuracy, there was a higher sensitivity in patients in a chronic stage of schizophrenia as compared to first-episode patients. These findings might result from more pronounced brain changes in older subjects with schizophrenia. In support of this previous studies report accelerated "brain aging" in schizophrenia potentially indicating a neurodegenerative process[33–35]. In addition this finding might result from secondary effects of the disease, not related to the underlying pathology but to environmental factors. Numerous studies report progressive brain changes to be associated with short-term[36] and long-term[37] antipsychotic treatment. Thus, pronounced brain changes and higher diagnostic accuracy of neuroimaging-based models in older patients might additionally result from long-standing antipsychotic treatment[35,38,39]. 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 a higher sensitivity, higher chlorpromazine equivalents was associated with higher specificity.

Comment [SJW4]: Yes, although I think this evidence is weak and it isn't supported by the BrainAGE paper

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[40,41]. Similarly there seem to be differences in brain changes between schizophrenic patients with predominantly positive and predominantly negative symptoms[42,43]. This might seem counterintuitive as previous studies indicate larger brain structural abnormalities in patients with predominantly negative symptom symptoms[42]. 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 to patients with negative symptoms and thus allows for greater diagnostic accuracy. It migth be hypothesized that patients with predominantely positive symptoms also received higher dosages of antipsychotic medication. Therefore the finding that positive symptoms are associated with higher diagnostic accuracy might be confounded by previous treatment. This is supported by the finding that chlorpromazine equivalent significantely affected specificity. Another more hypothetical explanation of the association between predominant positive symptoms and diagnostic accuracy might be related 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 full-blown psychotic patients as compared to patients with predominant 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 to the full-blown psychosis phenotype, which creates an area of diagnostic ambiguity for neuroimaging-based classification approaches. This would in

fact limit the maximal diagnostic accuracy that could be achieved by the neuroimaging-based predictive model.

Differences between multivariate methods

In our analysis there was substantial heterogeneity regarding the different statistical methods 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%). Both approaches showed almost identical sensitivity and specificity. Three studies[15–17] applied an artificial neural network model to structural MRI and PET data with slightly higher sensitivity (86 to 100%) and slightly higher specificity (85 to 100%). Two studies[13,14] 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 to other studies. However it needs to be noted that the comparison of different classification methods in the context of the present meta-analysis might be confounded by the characteristives of the investigated sample such as age, medication, symptoms and disease stage. Until today a systematic investigation of multiple classification algorythms is missing.

Limitations of the presented study

It needs to be noted that most of the published studies on neuroimaging-based predictive models for prediction of clinical diagnosis, 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 a 'proof of concept', showing that multivariate models are principally able to predict disease status. 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.[44], a detailed description of the patients' disease status, symptoms, length and course of illness, current medication or comorbidities is crucial information to evaluate the potential of such a model to enter clinical practice. Also as pointed out by the results in the present meta-analysis, clinical factors such as age or symptoms affect diagnostic accuracy. 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 in order to advance from a theoretical field of research to clinically applicable diagnostic methods, future studies should provide detailed information regarding their samples. Only in this way the applicability of multivariate methods in the clinic to various patient samples, subsamples or disease states can be evaluated.

It must be noted that in the studies included in the present analysis only compared patients with schizophrenia to healthy controls subjects. This of course represents an enormous simplification of the every-day clinical diagnostic process. Clinicians not only need to differentiate patients from healthy individuals but also different psychiatric diagnoses as well as diagnostic subtypes from each other. Only few published studies until today have investigated the potential of neuroimaging-based

predictive models to e.g. separate different patient groups. 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 adress this issue might be the delineation of more homogenous patient subgroups within and across disease boundaries[45] by means of unsupervised and semisupervised analysis methods. Apart from diagnosis, a central concern of clinicians is the prediction of functional outcome and treatment response. Future studies need to adress the question of how well neuroimaging-based predictive models generalize. In the studies included in the present analysis data has been aquired on one site using the same scanners and scanning sequences. Thus it is not clear if similar accuracy can be achieved for data from different sites or if a site-specific classifier would be needed. However, we believe that the present results are encouraging and the potential of neuroimaging data to assist diagnosis and to become a part of day-to-day clinical routine should be further investigated.

Acknowledgements

We would like to thank the authors of the included studies for providing additional information.

References

1. Howes OD, Kambeitz J, Kim E, et al. The Nature of Dopamine Dysfunction in Schizophrenia and What This Means for Treatment: Meta-analysis of Imaging Studies. *Arch Gen Psychiatry*. 2012.
2. Minzenberg MJ, Laird AR, Thelen S, Carter CS, Glahn DC. Meta-analysis of 41 functional neuroimaging studies of executive function in schizophrenia. *Arch Gen Psychiatry*. 2009;66(8):811–822.
3. Chan RCK, Di X, McAlonan GM, Gong Q-Y. Brain Anatomical Abnormalities in High-Risk Individuals, First-Episode, and Chronic Schizophrenia: An Activation Likelihood Estimation Meta-analysis of Illness Progression. *Schizophr Bull*. 2009.
4. Fornito A, Yücel M, Patti J, Wood SJ, Pantelis C. Mapping grey matter reductions in schizophrenia: an anatomical likelihood estimation analysis of voxel-based morphometry studies. *Schizophr Res*. 2009;108(1-3):104–113.
5. Honea R, Crow TJ, Passingham D, Mackay CE. Regional deficits in brain volume in schizophrenia: a meta-analysis of voxel-based morphometry studies. *Am J Psychiatry*. 2005;162(12):2233–2245.
6. Kapur S, Phillips AG, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatry*. 2012;17(12):1174–1179.
7. Borgwardt S, Radua J, Mechelli A, Fusar-Poli P. Why are psychiatric imaging methods clinically unreliable? Conclusions and practical guidelines for authors, editors and reviewers. *Behav Brain Funct BBF*. 2012;8:46.
8. Borgwardt S, Fusar-Poli P. Third-generation neuroimaging in early schizophrenia: translating research evidence into clinical utility. *Br J Psychiatry*. 2012;200(4):270–272.
9. Davatzikos C. Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *NeuroImage*. 2004;23(1):17–20.
10. McIntosh AR, Lobaugh NJ. Partial least squares analysis of neuroimaging data: applications and advances. *NeuroImage*. 2004;23 Suppl 1:S250–263.
11. Davatzikos C, Shen D, Gur RC, et al. Whole-brain morphometric study of schizophrenia revealing a spatially complex set of focal abnormalities. *Arch Gen Psychiatry*. 2005;62(11):1218–1227.
12. Kawasaki Y, Suzuki M, Kherif F, et al. Multivariate voxel-based morphometry successfully differentiates schizophrenia patients from healthy controls. *NeuroImage*. 2007;34(1):235–242.
13. Anderson A, Dinov ID, Sherin JE, Quintana J, Yuille AL, Cohen MS. Classification of spatially unaligned fMRI scans. *NeuroImage*. 2010;49(3):2509–2519.
14. Greenstein D, Malley JD, Weisinger B, Clasen L, Gogtay N. Using multivariate machine learning methods and structural MRI to classify childhood onset schizophrenia and healthy controls. *Front Psychiatry Front Res Found*. 2012;3:53.
15. Josin GM, Liddle PF. Neural network analysis of the pattern of functional connectivity between cerebral areas in schizophrenia. *Biol Cybern*. 2001;84(2):117–122.
16. Bose SK, Turkheimer FE, Howes OD, et al. Classification of schizophrenic patients and healthy controls using [18F] fluorodopa PET imaging. *Schizophr Res*. 2008;106(2-3):148–155.
17. Rathi Y, Malcolm J, Michailovich O, et al. Biomarkers for identifying first-episode schizophrenia patients using diffusion weighted imaging. *Med Image Comput Comput-Assist Interv MICCAI Int Conf Med Image Comput Comput-Assist Interv*. 2010;13(Pt 1):657–665.

18. Castellani U, Rossato E, Murino V, et al. Classification of schizophrenia using feature-based morphometry. *J Neural Transm Vienna Austria 1996*. 2012;119(3):395–404.
19. Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci*. 2009;12(5):535–540.
20. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009;339.
21. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529–536.
22. Gatsonis C, Paliwal P. Meta-analysis of diagnostic and screening test accuracy evaluations: methodologic primer. *AJR Am J Roentgenol*. 2006;187(2):271–281.
23. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982–990.
24. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med*. 2001;20(19):2865–2884.
25. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JAC. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostat Oxf Engl*. 2007;8(2):239–251.
26. Ioannidis J, Lau J. Evolution of treatment effects over time: empirical insight from recursive cumulative metaanalyses. *Proc Natl Acad Sci U S A*. 2001;98(3):831–836.
27. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol*. 2005;58(9):882–893.
28. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2013.
29. Doebler P. Meta-Analysis of Diagnostic Accuracy with mada. 2012.
30. Fan Y, Shen D, Gur RC, Gur RE, Davatzikos C. COMPARE: classification of morphological patterns using adaptive regional elements. *IEEE Trans Med Imaging*. 2007;26(1):93–105.
31. Liu M, Zeng L-L, Shen H, Liu Z, Hu D. Potential risk for healthy siblings to develop schizophrenia: evidence from pattern classification with whole-brain connectivity. *Neuroreport*. 2012;23(5):265–269.
32. Shen H, Wang L, Liu Y, Hu D. Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI. *NeuroImage*. 2010;49(4):3110–3121.
33. Kempton MJ, Stahl D, Williams SCR, DeLisi LE. Progressive lateral ventricular enlargement in schizophrenia: a meta-analysis of longitudinal MRI studies. *Schizophr Res*. 2010;120(1-3):54–62.
34. Andreasen NC, Nopoulos P, Magnotta V, Pierson R, Ziebell S, Ho B-C. Progressive brain change in schizophrenia: a prospective longitudinal study of first-episode schizophrenia. *Biol Psychiatry*. 2011;70(7):672–679.
35. Fusar-Poli P, Smieskova R, Kempton MJ, Ho BC, Andreasen NC, Borgwardt S. Progressive brain changes in schizophrenia related to antipsychotic treatment? A meta-analysis of longitudinal MRI studies. *Neurosci Biobehav Rev*. 2013;37(8):1680–1691.
36. Tost H, Braus DF, Hakimi S, et al. Acute D2 receptor blockade induces rapid,

reversible remodeling in human cortical-striatal circuits. *Nat Neurosci*. 2010;13(8):920–922.
37. Navari S, Dazzan P. Do antipsychotic drugs affect brain structure? A systematic and critical review of MRI findings. *Psychol Med*. 2009;39(11):1763–1777.
38. Smieskova R, Fusar-Poli P, Allen P, et al. The effects of antipsychotics on the brain: what have we learnt from structural imaging of schizophrenia?--a systematic review. *Curr Pharm Des*. 2009;15(22):2535–2549.
39. Ho B-C, Andreasen NC, Ziebell S, Pierson R, Magnotta V. Long-term antipsychotic treatment and brain volumes: a longitudinal study of first-episode schizophrenia. *Arch Gen Psychiatry*. 2011;68(2):128–137.
40. Modinos G, Costafreda SG, van Tol M-J, McGuire PK, Aleman A, Allen P. Neuroanatomy of auditory verbal hallucinations in schizophrenia: a quantitative meta-analysis of voxel-based morphometry studies. *Cortex J Devoted Study Nerv Syst Behav*. 2013;49(4):1046–1055.
41. Palaniyappan L, Balain V, Radua J, Liddle PF. Structural correlates of auditory hallucinations in schizophrenia: a meta-analysis. *Schizophr Res*. 2012;137(1-3):169–173.
42. Koutsouleris N, Gaser C, Jäger M, et al. Structural correlates of psychopathological symptom dimensions in schizophrenia: a voxel-based morphometric study. *NeuroImage*. 2008;39(4):1600–1612.
43. Nenadic I, Sauer H, Gaser C. Distinct pattern of brain structural deficits in subsyndromes of schizophrenia delineated by psychopathology. *NeuroImage*. 2010;49(2):1153–1160.
44. Devillé WL, Buntinx F, Bouter LM, et al. Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Med Res Methodol*. 2002;2(1):9.
45. Insel T, Cuthbert B, Garvey M, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry*. 2010;167(7):748–751.

Figure legends

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).

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).

Table 1: Results from bivariate meta-analyses applying the approach by Reitsma et al. (2005). Positive LR, negative LR and DOR are estimated via MCMC (Zwindermann & Bossuyt, 2008).

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.

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.

.