Schizophrenia Research xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Schizophrenia Research

journal homepage: www.elsevier.com/locate/schres

Improved individualized prediction of schizophrenia in subjects at familial high risk, based on neuroanatomical data, schizotypal and neurocognitive features

Eleni Zarogianni [a,*], Amos J. Storkey [b], Eve C. Johnstone [a], David G.C. Owens [a], Stephen M. Lawrie [a]

[a] Division of Psychiatry, School of Clinical Sciences, University of Edinburgh, The Royal Edinburgh Hospital, Morningside Park, UK
[b] Institute for Adaptive and Neural Computation, University of Edinburgh, UK

ARTICLE INFO

Article history:
Received 20 July 2016
Received in revised form 29 August 2016
Accepted 29 August 2016
Available online xxxx

Keywords:
MRI
Machine learning
Support vector machine
Recursive feature elimination
Schizophrenia
Prediction
Familial HR

ABSTRACT

To date, there are no reliable markers for predicting onset of schizophrenia in individuals at high risk (HR). Substantial promise is, however, shown by a variety of pattern classification approaches to neuroimaging data. Here, we examined the predictive accuracy of support vector machine (SVM) in later diagnosing schizophrenia, at a single-subject level, using a cohort of HR individuals drawn from multiply affected families and a combination of neuroanatomical, schizotypal and neurocognitive variables. Baseline structural magnetic resonance imaging (MRI), schizotypal and neurocognitive data from 17 HR subjects, who subsequently developed schizophrenia and a matched group of 17 HR subjects who did not make the transition, yet had psychotic symptoms, were included in the analysis. We employed recursive feature elimination (RFE), in a nested cross-validation scheme to identify the most significant predictors of disease transition and enhance diagnostic performance. Classification accuracy was 94% when a self-completed measure of schizotypy, a declarative memory test and structural MRI data were combined into a single learning algorithm; higher than when either quantitative measure was used alone. The discriminative neuroanatomical pattern involved gray matter volume differences in frontal, orbitofrontal and occipital lobe regions bilaterally as well as parts of the superior, medial temporal lobe and cerebellar regions. Our findings suggest that an early SVM-based prediction of schizophrenia is possible and can be improved by combining schizotypal and neurocognitive features with neuroanatomical variables. However, our predictive model needs to be tested by classifying a new, independent HR cohort in order to estimate its validity.

1. Introduction

Schizophrenia is a severe and debilitating psychiatric disorder, contributing significantly to the global burden of disease (Jablensky, 1997). The pathophysiological underpinnings of schizophrenia remain to date unclear, although it has been associated with structural alterations that span multiple brain regions (Wright et al., 2000, Shenton et al. 2001, Honea et al., 2005, Haijma et al., 2013) and neurocognitive abnormalities, affecting many areas of cognitive performance and functioning (Heinrichs and Zakzanis, 1998, Fusar-Poli et al., 2012). However, the exact onset and extent of these abnormalities remain to be elucidated.

Recently, neuroanatomical and neurocognitive studies of individuals considered at high-risk (HR) for developing schizophrenia have gained much attention as they promise to identify surrogate markers of vulnerability and transition (Borgwardt et al., 2011, Cooper et al., 2014, Bois et al., 2015, Fusar-Poli et al., 2012, Brewer et al., 2006). A host of structural MRI (sMRI) studies in HR cohorts, including voxel-based morphometry (VBM), have reported findings that although sometimes contradictory (Wood et al., 2005, Pantelis et al., 2009, Klauser et al., 2015) mainly involved neuroanatomical abnormalities in spatially distributed regions of the frontal, temporal and cingulate cortices (Fusar-Poli et al., 2011, Borgwardt et al., 2007a, Pantelis et al., 2003, Meisenzahl et al., 2008, Job et al., 2003), qualitatively similar to the established state albeit at a lesser extent.

The key clinical challenge lies, however, in identifying those HR subjects that are most likely to later develop schizophrenia and other related psychoses (Smieskova et al., 2010), as this could advise suitable intervention strategies, shown to lead to better clinical outcomes (Marshall & Lockwood 2004, Riecher-Rössler et al., 2006). Recent VBM studies found decreased gray matter (GM) volumes in the prefrontal, medial temporal, anterior cingulate and cerebellar cortices (Pantelis et al., 2003, Borgwardt et al., 2007a, Borgwardt et al., 2007b, Job et al., 2006, Borgwardt et al., 2008, Koutsouleris et al., 2009b, Mechelli et al., 2011) in HR individuals who subsequently developed psychosis

* Corresponding author at: Kennedy Tower, Division of Psychiatry, School of Clinical Sciences, University of Edinburgh, The Royal Edinburgh Hospital, Morningside Park, EH10 5HF, UK.
E-mail address: E.Zarogianni@sms.ed.ac.uk (E. Zarogianni).

http://dx.doi.org/10.1016/j.schres.2016.08.027
0920-9964/

compared to HR subjects with no transition. In the cognitive domain, HR subjects with later conversion showed widespread deficits in verbal abilities (fluency and memory) and memory functioning (Pukrop et al., 2007, Lencz et al., 2006, Johnstone et al., 2005, Erlenmeyer-Kimling et al., 2000, Whyte et al., 2006) than those who did not convert. In an effort to optimize prediction of psychosis transition, the integration of neurocognitive (Lencz et al., 2006, Riecher-Rössler et al., 2009) and clinical (Cannon et al., 2008, Ruhrmann et al., 2010, Demjaha et al. 2012) or neuropsychological variables (Pukrop et al., 2007, Seidman et al., 2010) into a single multivariate model was shown to significantly improve predictive accuracy, reaching up to 80% (Lencz et al., 2006, Riecher-Rössler et al., 2009).

The clinical utility of these findings is however limited as the analyses performed could only characterize group differences and do not permit inferences at an individual-level. This can be effectively addressed by machine-learning methods, which allow predictions that are specific to an individual and can additionally capture complex and subtle patterns of alteration (Klöppel et al., 2012, Zarogianni et al., 2013, Orrù et al., 2012). Therefore, there is a higher chance of eventually translating findings into clinical practice and producing reliable biomarkers.

The vast majority of sMRI- and machine learning-based studies published so far focused on the classification of schizophrenia against healthy control (HC) individuals (Davatzikos et al., 2005, Fan et al., 2008, Sun et al., 2009, Nieuwenhuis et al., 2012, Schnack et al., 2014), with only few studies contrasting the clinically useful groups of HR subjects with and without later conversion to psychosis. Within this context, Koutsouleris et al. (2009a) recently demonstrated that a spatial network of distributed GM volume abnormalities allows the prediction of transition to psychosis in a clinically-defined HR cohort, using a Support Vector Machine (SVM) model, with an accuracy of 82%. A subsequent investigation by the same group replicated this remarkable finding in an independent clinical HR population (84% accuracy, Koutsouleris et al., 2012a). A similarly high predictive accuracy was achieved when the same group implemented a SVM model using neurocognitive data from a pool of test batteries to distinguish between converters and non-converters (Koutsouleris et al., 2012b). However, these investigations concern heterogeneous groups of individuals, presented at health services with various attenuated and/or intermittent psychotic symptoms and it remains unclear whether similar results could be achieved in familial HR cohorts.

Here, the primary aim was to investigate the predictive capacity of SVM in later diagnosing schizophrenia in a HR cohort drawn from multiply affected families, derived from the Edinburgh High Risk Study (EHRS), using baseline sMRI data. As known, evidence for a high risk state emerges partly from having a family history of schizophrenia (Lawrie et al., 2001). Previous VBM investigations in EHRS revealed longitudinal changes in temporal lobe structures that provided good positive predictive power of future transition to schizophrenia (up to 60%, Job et al., 2006). Additionally, baseline measures of schizotypy and a neuropsychological test of verbal memory and learning were shown to be good predictors of transition, providing strong negative predictive power (up to 94%, Johnstone et al., 2005). We thus aimed to investigate whether inclusion of those variables into the neuroanatomical model would increase the predictive performance of our classifier. In this context, recent machine learning-based studies examining the diagnostic performance of combined sMRI data and neuropsychological (Karageorgiou et al., 2011), cognitive (Pettersson-Yeo et al., 2013) or genetic information (Yang et al., 2010) reported increased classification performance (Pettersson-Yeo et al., 2014) in distinguishing between schizophrenia patients and HC or HR and first-episode individuals.

In light of all these, we hypothesized that the application of SVM to baseline sMRI data would allow prediction of future development of schizophrenia in our familial HR cohort and that a combination of sMRI with schizotypal and neurocognitive data would enhance predictive performance compared to either individual measures alone.

2. Materials and methods

2.1. Subjects

The subject material was gathered as part of the EHRS; a longitudinal study that recruited young people from multiply affected families with schizophrenia. The recruitment process has been detailed previously (Hodges et al., 1999, Johnstone et al., 2000). Briefly, 160 individuals aged 16–25 years, with no previous history of psychiatric problems, were drawn throughout Scotland and were identified as HR on the basis of having two or more relatives affected with schizophrenia. Subjects were followed up for 10 years during which they underwent a series of clinical, behavioural and neuroimaging assessments every 18 months.

Psychopathology was assessed at entry and follow-up by the Present State Examination (PSE) (Wing et al., 1974) and allowed the classification of subjects based on the presence of (psychotic) symptoms (Table 1, Johnstone et al., 2000) and/or diagnosis of schizophrenia, which was further validated by the ICD-10 (WHO, 1994).

All subjects were antipsychotic-naïve at baseline and at follow-up or until they were clinically diagnosed with schizophrenia. Subjects did not receive any other medication, such as anti-depressants, at study-entry. From those HR subjects who provided complete data and had a MRI scan, 17 were later diagnosed with schizophrenia (after an average

Table 1
Baseline sociodemographic and behavioural assessment variables of the groups.

	HR[*ill*]	HR[symp]	*P*	HR[symp]$_{test}$
Number of participants	17	17		40
Mean age at baseline (SD)	20.07(2.37)	20.03(2.6)	ns[a]	21.07(4.59)
Male (%)	11(64.7)	11(64.7)	ns[b]	14(35)
Mean RISC score (SD)	39.88(10.6)	25.23(11.75)	<0.01[a]	33.45(13.39)
Mean RAVLT, trials 1–5 (SD)	47.64(7.49)	53.41(7.45)	ns[a,c]	51.15(11.59)
Mean WAIS-IQ (SD)	98.64(12.93)	98.98(14.7)	ns[a]	96.42(12.64)
Handedness			ns[b]	
Right	16	15		34
Left	0	1		4
Mixed	1	1		2
Social class of origin			ns[b]	
I and II	2	3		7
III and IV	13	10		22
V and VI	2	3		10
unclassifiable	0	1		1
Cannabis use at baseline			ns[b]	
None	8	12		26
Occasional	5	4		10
Frequent	4	1		4
Smoking cigarettes			ns[b]	
None	7	9		21
<10	6	4		12
10–20	3	2		4
>20	1	2		3
Symptoms severity (PSE rating)			ns[b]	
No symptoms	2	7		13
Neurotic symptoms only	4	3		13
Partially held psychotic symptoms	9	6		13
Isolated and/or transient psychotic symptoms	2	1		1

HR[*ill*]: individuals at high familial risk who developed schizophrenia during follow-up period; HR[symp]: individuals at high familial risk who remained well but developed psychotic symptoms during follow-up period; IQ, Intelligence Quotient; RISC, Rust Inventory of Schizotypal Cognitions; RAVLT, Rey Auditory Verbal Learning Test; WAIS-R, Wechsler Adult Intelligence Scale- Revised. Social class of origin was based on the father's occupation at the time of the subject's birth using the Occupational Classification of the Registrar General (HMSO, 1991).

Statistical analyses are shown for the HR[*ill*]-HR[symp] contrast.

[a] All other variables were evaluated using *t*-tests, P significance.

[b] Fisher's exact tests were applied to sex, handedness, social class of origin, present cannabis use, smoking and symptom severity scores.

[c] Effect size was r = 0.54.

Please cite this article as: Zarogianni, E., et al., Improved individualized prediction of schizophrenia in subjects at familial high risk, based on neuroanatomical data, schizotypal ..., Schizophr. Res. (2016), http://dx.doi.org/10.1016/j.schres.2016.08.027