Citation for published version (APA):
Pattern of Activation during Delayed Matching to Sample Task predicts Functional Outcome in People at Ultra High Risk for Psychosis

Irina Falkenberg¹,²*, M.D.; Isabel Valli¹, M.D., Ph.D.; Marie Raffin¹,³, M.D., Ph.D.; Matthew R. Broome⁴,⁵, M.D., Ph.D.; Paolo Fusar-Poli¹, M.D., Ph.D.; Pall Matthiasson¹, M.D., Ph.D.; Marco Picchioni¹, M.D., Ph.D.; Philip McGuire¹, F.R.C.Psych., Ph.D.

¹ Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, London, UK
² Department of Psychiatry and Psychotherapy, Philipps-University Marburg, Marburg, Germany
³ Department of Child and Adolescent Psychiatry, Université Pierre et Marie Curie, Hôpital Pitié-Salpêtrière, Paris, France
⁴ Department of Psychiatry, University of Oxford, Oxford, UK.
⁵ Oxford Health NHS Foundation Trust, Warneford Hospital, Oxford, UK.

* Corresponding Author:
Irina Falkenberg, Department of Psychiatry and Psychotherapy, Philipps-University, Rudolf-Bultmann-Str. 8, 35039 Marburg, Germany. Phone: +49 (0) 6421-5865200; e-mail: Irina.Falkenberg@med.uni-marburg.de

Short title: Prediction of functional outcome in psychosis risk
Abstract

**Background:** Clinical outcomes in people identified as at ultra-high risk (UHR) for psychosis are remarkably heterogeneous, and are difficult to predict on the basis of the presenting clinical features. Individuals at UHR are at risk of poor functional outcome regardless of development of psychotic disorder. The aim of the present study was to assess whether there is a relationship between functional neuroimaging measures at presentation and functional outcome as measured by the GAF three years after scanning.

**Methods:** Functional magnetic resonance imaging (fMRI) data were collected during an object working memory task in 34 ultra-high risk (UHR) subjects and 20 healthy controls. On the basis of their GAF scores at follow up, the UHR participants were divided into subgroups with good and poor functional outcomes, respectively.

**Results:** At baseline, the UHR group differed from controls in showing altered frontal and cuneus/posterior cingulate activation. Significant group x task interactions were found in the left cuneus and posterior cingulate gyrus, reflecting differential responses to the task conditions. Within the UHR sample, the subgroup with a poor functional outcome exhibited altered activation in frontal, temporal and striatal regions, and reduced deactivation within default-mode network regions, relative to those with a good outcome. Within the whole UHR sample, in these regions the local task response was correlated with the GAF score at follow up.

**Conclusions:** The findings suggest a potential role of functional neuroimaging in the prediction of outcomes in people at high clinical risk of psychosis.

**Key words:** psychosis; working memory; functional outcome; UHR; default mode network; fMRI
1. Introduction

A substantial proportion of individuals at Ultra High Risk for psychosis (UHR) will develop a psychotic disorder within 3 years (Fusar-Poli et al., 2012). The onset of psychosis has been the predominant outcome in prospective studies of UHR subjects, including those using neuroimaging. Transition to psychosis is operationally defined in terms of the severity of positive psychotic symptoms (Fusar-Poli et al., 2008). However, the threshold level for these symptoms is somewhat arbitrary (Lin et al., 2015), and other clinical features such as the severity of negative symptoms and the subject’s level of functioning are not taken into account (Lin et al., 2015). As a result, individuals who have not met criteria for transition (in terms of positive symptoms) can still have a very poor clinical outcome, because they have prominent negative symptoms and/or a low level of functioning (for example as measured by the GAF; Lin et al., 2015). Furthermore, data from studies in schizophrenia suggest that in terms of recovery, patients often regard their overall level of functioning as more important for their quality of life than the severity of their positive symptoms (Cichocki et al., 2015).

Working memory impairment is a cardinal cognitive feature of psychotic disorders (Forbes et al., 2009) and is evident in individuals at UHR (P. Fusar-Poli et al., 2012a). In both patients with schizophrenia and individuals at UHR, working memory impairments have been linked to poor functional outcome (Goghi et al., 2014; Vesterager et al., 2012). Functional MRI studies of tasks that engage working memory have identified differences in regional activation between UHR subjects and controls (e.g. Broome et al., 2010). Moreover, a follow up study suggested that longitudinal improvements in level of functioning were related to changes in activation during a working memory task, although the sample size was small (Fusar-Poli et al., 2010).

The aim of the present study was to determine whether functional neuroimaging data collected during a working memory task could predict functional outcome in people at UHR for psychosis, irrespective of transition to the full-blown disorder. Level of functioning was measured with the DSM-IV GAF-scale, a widely used instrument for assessing overall psychological, social and
occupational functioning (Aas, 2010; Jones et al., 1995). The GAF is well-established, highly generalizable and has been used in many outcome studies including those of people at UHR (Brandizzi et al., n.d.; Kambeitz-Ilankovic et al., 2015; Lin et al., 2015). We used the delayed matching to sample (DMTS) task from the Cambridge Neuropsychological Test Automated Battery (CANTAB; Robbins et al., 1994), which permits the experimental manipulation of task demand. This is potentially useful in studies of UHR subjects, as there is some evidence that differential activation in this group becomes increasingly evident as the demands on working memory are increased (Broome et al., 2010; Fryer et al., 2013).

On the basis of data from previous neuroimaging studies of working memory in this group (Broome et al., 2010) we first predicted that UHR individuals would show altered prefrontal and parietal activation relative to controls, and that these differences would vary with increasing task demand. We then tested the hypothesis that within the UHR sample, the pattern of activation at baseline would be related to the level of functional outcome at follow up, as defined using the Global Assessment of Functioning (GAF) scale.
2 Methods and Materials

2.1 Participants

The study was approved by the South London and Maudsley research ethics committee. All subjects gave written informed consent after a full description of the study. The sample consisted of 34 subjects at ultra-high risk (UHR) and 20 healthy controls (CTRL), all right handed and without any personal or family history of psychiatric disorders. Substance misuse disorder, an IQ<70, history of another CNS disorder or head injury with documented cognitive sequelae, were exclusion criteria for all groups. Demographic and clinical details are provided in Table 1.

2.1.1 Ultra-high risk group (UHR)

Individuals meeting Personal Assessment and Crisis Evaluation (PACE) criteria for the UHR state of psychosis (Yung et al., 1998) were recruited from Outreach and Support in South London (OASIS; Broome et al., 2005) clinical service. The diagnosis was based on assessment by two experienced clinicians using the Comprehensive Assessment for the At-Risk-Mental State (CAARMS; Yung et al., 2005; Table 1) and a consensus meeting with the clinical team (see Supplementary Material).

2.1.2 Control group (CTRL)

Healthy participants were recruited via advertisements in local media. All individuals lived in the same London borough as the clinical participants. The groups differed in terms of age and IQ estimates (Table 1), so these were included as covariates in all subsequent analyses.

Table 1 to be placed about here
2.2 Clinical and neuropsychological / IQ assessment

Prior to scanning, all subjects were interviewed using the SCID-1 and SCID-2 interviews. Of 34 UHR subjects 9 (26%) had depression and/or anxiety. Family history was assessed using the Family Interview for Genetic Studies (FIGS). The Positive and Negative Syndrome Scale (PANSS) rated symptom severity (Table 1). Handedness was assessed with the Edinburgh Handedness Inventory and pre-morbid Wechsler Adult Intelligence Scale-Revised IQ (WAIS-R) using the National Adult Reading Test (NART) and standard tables.

2.3 Outcome measures

Based on the median GAF score of 70 at follow-up (cf. Kambeitz-Ilanvodic et al., 2015), UHR participants were subdivided into good (UHRg; N=10, GAF > 70) and poor functional outcome groups (UHRp; N=9, GAF ≤ 70).

2.4 Task

Functional MRI data were acquired while subjects performed a modified version of the CANTAB delayed matching to sample (DMTS) task (Picchioni et al., 2007). The DMTS paradigm requires encoding and maintaining a complex visual pattern that, after a variable delay period, must be discriminated from a set with three additional distracters. Trials during which the recognition phase immediately follows the stimulus presentation are compared to trials testing recognition after a longer delay. Simultaneous and delayed task conditions are randomized, with all conditions being identical until after stimulus presentation. This design ensures comparable encoding for all task conditions, and restricts the difference between task conditions to the ability to hold information online. It is thus assumed that increased errors in the delayed compared to the immediate matching condition reflect processes that are not attentional or perceptual but more closely relate to WM maintenance. Stimuli, each subtending an angle of 5°, were presented using Visual Basic (Microsoft, Redmond) on a black screen, viewed through a mirror. Subjects initially focused on a central fixation cross. Each trial consisted of four phases (See Supplementary Material). During initial ‘encoding’,...
subjects were presented with a complex abstract pattern (the sample) for 5 s in the centre of the screen. The sample consisted of a rectangular target divided into quadrants; each differed in shape and colour. The next ‘maintenance’ phase involved a variable length delay (simultaneous=0 s, 4 or 12 s) during which the subjects were instructed to memorize the sample while maintaining fixation on the central cross. In the third ‘recognition’ phase, subjects were shown four patterns in a North, South, East and West distribution around the central location for 6 s and asked to identify the sample by pressing a joystick in the corresponding direction with their right hand. One pattern was identical to the sample, one a novel distractor (D-Error), one the same colour but with a different shape distribution (S-Error), and the other the same shape but with different colour distribution to the original sample (C-Error). To discourage the use of mnemonic strategies based on encoding a single quadrant, all four choice patterns shared one random quadrant in common with the sample. The final phase of each trial involved a delay during which the subjects again focused on a fixation cross which was designed to equalize the inter-trial (encoding) interval to 27 s, while randomly varying the inter-stimulus (recognition) interval, with the length of the delay dependent upon the duration of the preceding maintenance phase. The task comprised 42 trials, 14 for each maintenance delay, presented in a pseudo-random order in two runs of approximately 10 min.

2.5 fMRI image acquisition

Gradient echoplanar imaging (EPI) data were acquired on a GE Signa 1.5 T system (General Electric, Milwaukee). A quadrature birdcage head coil was used for RF transmission and reception. 324 T2*-weighted images depicting BOLD contrast were acquired over 10 min (for each run) at each of 22 near-axial non-contiguous 7-mm thick planes parallel to the intercommissural (AC-PC) line: TE 40 ms, TR 2 s, in-plane resolution 7 mm, interslice gap 0.7 mm. This EPI dataset provided almost complete brain coverage. The first four images were discarded to allow the magnetization to reach equilibrium amplitude. A jittered acquisition sequence optimised sampling of the BOLD response. Individual brain activation maps were co-registered to a ‘whole-head’ gradient echo image of superior spatial
resolution acquired on each subject. This structural scan had the following acquisition parameters: TE 40 ms, TR 3 s, 43 slices, in-plane resolution 3 mm, interslice gap 0.3 mm.

2.6 Behavioural data

Response accuracy and latency were recorded on a personal computer using Visual Basic (Microsoft) and analysed in SPSS Version 18.0 (SPSS Inc. Released 2009. PASW Statistics for Windows, Version 18.0. Chicago: SPSS Inc).

2.7 Follow up procedure

All UHR subjects were contacted and invited to participate in a set of follow up assessments comprising re-assessment of CAARMS and GAF, assessment of vocational status (categories included: (1)=student, (2)=part-time paid, (3)=part-time unpaid, (4) full-time paid, (5) full-time unpaid, (6)=unemployed), satisfaction with social life as perceived by the client (an 11-point Likert-scale (0=not at all satisfied, 10=extremely satisfied)). No re-assessment was carried out in 4 UHR participants (11.8 %) who had developed psychosis. Eleven subjects were uncontactable. Re-assessments were completed in 19 of the original sample. The mean interval between scanning and follow up was 3.58 ± 2.2 years. There was a significant difference in IQ between the UHR subjects who completed the re-assessment relative to those who did not, with the mean IQ being lower in the completers (mean IQ 94.7 (SD 11.9) vs 105.4 (SD 10.2), T=-2.77, p=0.009; cf. Table S1), which was then controlled for in the subsequent analysis.

2.8 Data analysis

2.8.1 Individual analysis

Data were analyzed with XBAM v4.1 (Institute of Psychiatry, London, UK).
2.8.2 Group analysis

Between group comparisons (UHR vs CTRL and UHRg vs UHRp) were performed using nonparametric ANCOVA with a voxelwise threshold of $p=.05$ and the clusterwise threshold set such that the total number of false positive clusters per brain volume was less than 1. The $p$ value for each respective analysis is reported.

2.8.3 Analysis of behavioural data

A repeated measures ANCOVA was performed to test for differences in task performance between groups. All $p$-values were Bonferroni corrected. In cases where the data did not meet the assumption of sphericity, a Greenhouse-Geisser correction was applied.
3. Results

3.1 Behavioural results

3.1.1 UHR vs CTRL

The main effects of group (F = .18, df = 1, p = .67), task (F =2.62, df = 3, p = .8) and the group by task interaction (F = 1.4, df = 2, p = .25) on response accuracy were all non-significant.

3.2 Outcome measures

3.2.1 Level of Functioning

Table 2 summarizes the characteristics of the UHRg and UHRp outcome groups at baseline. In the UHRg subgroup the mean GAF score increased from 60.1 at baseline to 80.2 at follow up, with most individuals showing a net increase over the follow up period (Figure 1a). In the UHRp group, the mean GAF score decreased slightly from 58.8 to 57.8. In most subjects in this group there was little change in the GAF score over the follow up period (Figure 1b).

Figure 1 to be placed about here

3.3 Clinical status

At follow-up, 1 (10%) of those in the UHRg and 3 (33.3%) of those in the UHRp subgroup still fulfilled UHR criteria (X²=1.55, df=1, p=.21). The remaining subjects were operationally defined as being in remission from the UHR state. None of the UHR participants in either of the functional outcome subgroups had developed a psychotic disorder. The UHRg and UHRp groups did not differ in terms of interventions (CBT, medication) they had received during the 2 year period at OASIS (see Supplementary Results).
3.4 Vocational status

40% of the participants in the UHRg subgroup were in education or paid employment at follow-up, compared to 11% of those in the UHRp group, although this was not statistically significant ($X^2=6.3$, df=3, $p=.1$).

3.5 Social life

At follow up, there was a trend ($t=2.19$, $p=.065$) for participants in the UHRg subgroup to give higher ratings of satisfaction with their social life (mean rating 6.25; SD=1.25), than those in the UHRp group (3.8; SD=1.92).

Table 2 to be placed about here

3.6 Neuroimaging results

3.6.1 UHR vs CTRL

There were significant main effects of group (independent of task condition) in the right inferior frontal gyrus, left posterior cingulate, and the cuneus and cerebellum bilaterally. In the right inferior frontal gyrus, the UHR group showed greater activation than controls, reflecting activation in the UHR group across all 3 conditions, but deactivation in the controls (Figure S3). Conversely, in the right cuneus and left posterior cingulate gyrus, controls showed an increase in activation across all conditions, whereas in the UHR subjects there was deactivation (Table 3, Figure S3).

In many of these regions there were significant group x task interactions, reflecting differential responses to the task conditions (Table 3). For example, in the left cuneus and posterior cingulate gyrus there was a progressive increase in activation across the 3 conditions in the control group, whereas in UHR subjects there was deactivation which progressively diminished across the 3 conditions (Table 3, Figure S4).
3.6.2 Differences in activation: UHRg vs UHRp

The topographical distribution of group differences within the UHR sample was strikingly different to that for the UHR versus control comparison, with differences that were predominantly in frontal, temporal and anterior cingulate cortex, as opposed to the posterior cingulate cortex, cuneus and cerebellum. There were significant effects of group (independent of task condition) in the lateral and medial frontal cortex, the anterior cingulate and superior temporal cortex, and the putamen (Table 3, Figure 2). In the right middle frontal gyrus, the UHRg group showed activation across all conditions, whereas the UHRp group showed deactivation. In the right superior temporal gyrus both groups showed deactivation, but this was more marked in the UHRg than the UHRp group. In the left medial frontal and left anterior cingulate gyri and in the right putamen there was deactivation in the UHRg group but activation in the UHRp group (Table 3, Figure 2). No significant group x task interactions were found. GAF scores at follow-up were significantly correlated with baseline activation in the right middle frontal gyrus ($r_S = 0.64$, $p<0.001$), right superior temporal gyrus ($r_S = -0.44$, $p=0.001$), left medial frontal gyrus ($r_S = -0.54$, $p<0.001$), left anterior cingulate cortex ($r_S = -0.58$, $p<0.001$) and right putamen ($r_S = -0.46$, $p<0.001$; all Bonferroni-corrected).

Table 3 to be placed about here

Figure 2 to be placed about here
4. Discussion

Confirming our first hypothesis, we found that, independent of functional outcome, UHR subjects showed a differential pattern of regional activation during working memory relative to controls, consistent with previous research (Broome et al., 2010; Fusar-Poli et al., 2010). There was also evidence that many of these differences in activation were related to task demands, which varied across conditions. There were no group differences in task performance, indicating that the functional imaging differences are not simply attributable to the UHR participants performing the task at a poorer level than the controls.

The UHR group showed greater activation than controls in the right IFG. This finding is consistent with existing data of greater prefrontal engagement during WM across a broad range of schizophrenia and schizophrenia risk states. These include patients with chronic schizophrenia (Kim et al., 2010), first-episode psychosis patients (Fusar-Poli et al., 2007), people at UHR of psychosis (Fusar-Poli, 2012), unaffected siblings of patients with schizophrenia (Callicott et al., 2003) and finally healthy carriers of a Dysbindin schizophrenia susceptibility variant (Markov et al., 2009). Engagement of the IFG during DMTS is thought to be related to information encoding and retrieval and executive processes (Fransén, 2005; Picchioni et al., 2007). Greater activation in this area during performance of WM tasks has previously been interpreted as an indicator of “inefficient” cognitive processing (e.g. Callicott et al., 2003; Delawalla et al., 2008). In line with this hypothesis, IFG hyperactivity may be the substrate required to maintain normal task performance in our UHR subjects. In psychosis, prefrontal hyperactivation during WM is independent of illness duration or effects of medication (Fusar-Poli et al., 2007). Thus it may be that functional activity in this region is emerging as a core neurofunctional abnormality underlying an increased vulnerability to psychosis (Fusar-Poli, 2012).

The UHR group also exhibited greater deactivation than controls in the posterior cingulate cortex (PCC), a major node in the default mode network (DMN; Raichle et al., 2001). The DMN is a set of
brain regions active at rest and deactivated as subjects shift their attention to external, goal-oriented behavior. It has been suggested that the PCC has a central role in supporting internally directed cognition (Raichle et al., 2001). One interpretation of this result is that individuals at UHR may require greater DMN suppression in order to maintain task performance at the same level as healthy subjects.

Alterations in the DMN have previously been detected in functional neuroimaging studies of other UHR samples. Consistent with our results Yaakub et al. (Yaakub et al., 2013) observed increased task-related deactivation in the PCC during a working memory task in UHR subjects relative to controls. Fryer et al. (Fryer et al., 2013) found that when the demands of a working memory task were high, UHR subjects and patients with first-episode psychosis showed less deactivation of DMN regions such as the medial prefrontal cortex than controls. Contrary to our results, this study found no significant group effect in the PCC, which might be related to differences in paradigms, sample sizes and differences in analysis type (Fryer et al. used a ROI-based analysis while we used a whole brain approach). Using resting state data, Wotruba et al. (Wotruba et al., 2014) reported that UHR subjects showed altered connectivity between the right anterior insula and the PCC.

Our second prediction was that in UHR subjects, the pattern of brain activation at clinical presentation would be related to their level of functioning on average three years later. This hypothesis was confirmed, with differential engagement of frontal, anterior cingulate, superior temporal and striatal regions in subgroups defined according to their GAF score at follow up. Moreover, across the entire follow up sample, there was a significant correlation between the magnitude of activation in these regions and the GAF score at follow up. The topographical distribution of these differences was strikingly different to that of the differences in activation between the total UHR sample and the controls, which mainly involved parietal, occipital and cerebellar regions. During DMTS, the parietal cortex is thought to provide the spatial information required for directing attention to the salient stimulus in a complex scene (Constantinidis and Steinmetz, 2001; Picchioni et al., 2007), occipital regions are relevant to visual processing of stimul
and target selection (Picchioni et al., 2007) and cerebellar regions are involved in visual search and attention strategies (Picchioni et al., 2007; Yantis et al., 2002).

Fronto-temporal and fronto-striatal circuits are thought to play a major role in cognitive impairments in schizophrenia (Pantelis et al., 1997) and alterations in these circuits have been identified in neuroimaging studies of UHR subjects (Fusar-Poli et al., 2011a). However, most of these studies did not examine how these findings related to outcomes, although one reported that a poor functional outcome was linked to increased left IFG activation (Allen et al., 2014). In the present study, we also found that UHR subjects with a poor functional outcome failed to deactivate the ACC and medial frontal gyrus. These regions are part of the DMN, and have been implicated in self-referential cognitive processes and social cognition (Gusnard et al., 2001; Taylor et al., 2011), both of which are impaired in UHR subjects (Thompson et al., 2011), and have been linked to poor outcomes in this group (Thompson et al., 2011). One might therefore speculate that reduced activation in these regions may be associated with more self-reflective thinking and reduced ability to decode socially relevant cues, thus leading to lower psychosocial functioning.

Within our UHR sample, there was considerable heterogeneity in terms of functional outcome. In about half of the sample for whom follow-up data was available (UHRg), there was a substantial increase in the mean GAF score, rising from 60.1 at baseline to 80.2 at follow up (Figure 1a). In contrast, in the other half of the sample (UHRp), the mean GAF score did not improve, but remained at approximately the same low level (58.8 at baseline, and 57.8 at follow up; Figure 1b). These observations are consistent with previous findings (Allen et al., 2014), and highlight the importance of finding biomarkers that could be used to stratify UHR samples according to future clinical need (Reilly and McGuire, 2013). A large proportion (26.5%) of the overall UHR sample had a GAF score below 60 at follow up, which corresponds to a significant degree of functional impairment. This low level of functioning was despite them not having developed a psychotic disorder, supporting the notion that UHR subjects can have a poor clinical outcome without having made a transition to psychosis (Lin et al., 2015). Functional outcome in mental health disorders may be a better guide to
the level of clinical need and the costs of clinical care than traditional diagnostic categories (Gustavsson et al., 2011), and this may also be true in the case of outcomes in UHR subjects (Lin et al., 2015).

There are some limitations to the present study. The GAF measure of functional outcome could be confounded by psychiatric symptom severity, and therefore limited in its ability to shed light on functioning that is specifically associated with the risk of developing psychosis (Cornblatt et al., 2007). Functional consequences of being at high-risk of developing psychosis is a different construct from symptom severity and a new scale has recently been developed to assess these aspects separately (Cornblatt et al., 2012, 2007), but unfortunately wasn’t available when this study was conducted. We used a categorical approach to divide UHR participants into good and poor outcome groups based on their GAF score at follow-up. While the median-split method has been previously used with the same cut-off score of 70 in a neuroimaging study predicting functional outcome in an UHR sample (Kambeitz-Illankovic et al., 2015), this cut-off is not based on operationalized criteria for functional recovery in the UHR state, which, to date, are not available. We do, however, believe that the threshold of 70 is sensible both from a clinical point of view and with regards to Andreasens’ suggestion that complete recovery “implies the ability to function in the community, socially and vocationally, as well as being relatively free of disease-related psychopathology” (Andreasen et al., 2005). A GAF score in the 61-70 range corresponds to "some mild symptoms and some difficulty in social, occupational or school functioning", whereas scores from 71-80 correspond to "transient symptoms and no more than slight impairment in social, occupational or school functioning". We do, however, acknowledge that such binary decisions may be associated with some loss of information.

A large proportion of people at UHR for psychosis present to clinical services with a range of psychopathological symptoms other than attenuated psychotic symptoms. A recent meta-analysis in 1683 UHR subjects confirmed that baseline prevalence of co-morbid depressive and anxiety disorders is 41 % and 15 % respectively (P. Fusar-Poli et al., 2012b), with comorbidity rates varying greatly between studies (depression: 95% CI 32.5%–49.4%; anxiety: 95% CI 8.9%–25%; (P. Fusar-Poli
et al., 2012b)). It is thus surprising that the rate of comorbid depression and/or anxiety in our sample was relatively low (26%). This may in part be due to the fact that MRI studies in general may be particularly challenging for patients and, as comorbid diagnoses of anxiety or depression have been found to be associated with more severe psychopathology (P. Fusar-Poli et al., 2012b), we cannot exclude the possibility that subjects with comorbid affective disorders may have been less likely to consent to participate.

Although the total UHR sample at baseline was relatively large for a functional neuroimaging study, the number of subjects in the good and poor outcome subgroups were relatively small, partly because of subject attrition in the three years of follow up, which has to be considered a limitation of the present study. It is difficult for a single centre to recruit and scan a large sample of UHR subjects, but this issue can be addressed through multi-centre neuroimaging studies, and a number of these are ongoing. Most of the UHR participants were naïve to antipsychotic medication at the time of scanning, and the good and poor functional outcome subgroups did not differ in terms of clinical care they received from baseline to follow up so it is unlikely that the findings were confounded by effects of medication or CBT.

Our results provide further evidence that baseline fMRI measures can separate subgroups of UHR subjects according to outcomes at follow up (Allen et al., 2014; Fusar-Poli et al., 2011b), and extend this literature by illustrating that this applies to functional outcome, as well as the traditional outcome of transition to psychosis. A relationship between neuroimaging measures and functional outcomes is consistent with data from studies in patients with first episode psychosis (Lappin et al., 2014), and in chronic schizophrenia (Mitelman et al., 2005). More recently, functional outcome in UHR subjects has been found to be linked to the level of activation in inferior frontal cortex, the hippocampus, putamen and thalamus (Allen et al., 2014), and to thalamic glutamate levels (Allen et al., 2014). These observations suggest that neuroimaging has the potential to facilitate the prediction of outcomes, which is difficult to do on the basis of clinical measures alone. However, all of these findings have been at a group level: prediction of outcomes in clinical practice needs to be done on
the basis of neuroimaging data from a single individual. Addressing this issue is a key challenge for neuroimaging in this area. One approach that may help overcome this is the application of machine learning algorithms, which has recently been employed to facilitate prediction of transition to psychosis in UHR samples using MRI data (Koutsouleris et al., 2009). The same approach can be applied to any modality of imaging data, and to other measures of outcome, such as level of functioning (Reilly and McGuire, 2013).
References


Figure legends

**Figure 1** GAF score at baseline and at follow-up for each UHR subject in the good (a) and poor (b) outcome subgroups.

**Figure 2**

**A:** Brain areas in which UHRg and UHRp subjects showed differential activation. The right side of the picture corresponds to the right side of the brain.

**B:** Plots of mean parameter estimates in each group from foci in the left anterior cingulate (-11 -11 37), left medial frontal gyrus (-4 0 59), right middle frontal gyrus (47 26 31), right putamen (32 -19 -2) and the right superior temporal gyrus (58 -15 -2).