UNIVERSITYOF BIRMINGHAM

Heterogeneity and classification of recent onset psychosis and depression

Lalousis, Paris Alexandros; Wood, Stephen; Schmaal, Lianne; Chisholm, Katie; Griffiths, Lowri; Reniers, Renate; Bertolino, Alessandro; Borgwardt, Stefan; Brambilla, Paolo; Kambeitz, Joseph ; Lencer, Rebekka; Pantelis, Christos; Ruhrmann, Stephan; Salokangas, Raimo K.R.; Schultze-Lutter, Frauke; Bonivento, C.; Dwyer, Dominic; Ferro, Adele; Haidl, Theresa; Rosen, Marlene

Document Version
Peer reviewed version

[Link to publication on Research at Birmingham portal](#)

Draft Manuscript for Review. Please review online at http://mc.manuscriptcentral.com/oup/szbltn

Heterogeneity and Classification of Recent Onset Psychosis and Depression: a Multimodal Machine Learning Approach

SCHOLARONE™
Manuscripts

1 **Heterogeneity and Classification of Recent Onset Psychosis and Depression: a**

2 **Multimodal Machine Learning Approach**

3

4 Paris Alexandros Lalousis, MSc[1,2]; Stephen J. Wood, PhD[1,3,4]; Lianne Schmaal, PhD[3,4];

5 Katharine Chisholm, PhD[1,5]; Sian Lowri Griffiths, PhD[1,2]; Renate L.E.P Reniers, PhD[1,2,6];

6 Alessandro Bertolino, MD[7]; Stefan Borgwardt, MD[8]; Paolo Brambilla, MD[9,10]; Joseph

7 Kambeitz, MD[11]; Rebekka Lencer, MD, PhD[12]; Christos Pantelis, MB BS, MD, MRCPsych,

8 FRANZCP[13]; Stephan Ruhrmann, MD[14]; Raimo K. R. Salokangas, MD, PhD, MSc[15]; Frauke

9 Schultze-Lutter, PhD[16]; Carolina Bonivento, PhD[17]; Dominic Dwyer, PhD[11]; Adele Ferro,

10 PsyD, PhD[10]; Theresa Haidl, MD[14]; Marlene Rosen, MSc[14]; Andre Schmidt, PhD[8]; Eva

11 Meisenzahl, MD[16]; Nikolaos Koutsouleris, MD[11*]; Rachel Upthegrove, MBBS FRCPsych,

12 PhD[1,2,18*] and the PRONIA Consortium#

13 * Joint Senior Authors

14 #See PRONIA consortium author list in the supplement

15

16 **Author Affiliations:**

17 1 -Institute for Mental Health, University of Birmingham, Birmingham, United Kingdom

18 2 -Centre for Human Brain Health, University of Birmingham, Birmingham, United

19 Kingdom

20 3 -Orygen, the National Centre of Excellence in Youth Mental Health; Melbourne, Australia

21 4 -Centre for Youth Mental Health, The University of Melbourne, Parkville, Australia

22 5 -Department of Psychology, Aston University, United Kingdom

23 6 -Institute of Clinical Sciences, University of Birmingham, United Kingdom

24 7 -Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of

25 Bari Aldo Moro, Bari, Italy

26 8 -Department of Psychiatry, University of Basel, Basel, Switzerland

27 9 -Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy

28 10 - Department of Neurosciences and Mental Health, IRCCS Ca' Granda Ospedale

29 Maggiore Policlinico, Milan, Italy.

30   11 -Department of Psychiatry and Psychotherapy, Ludwig Maxmilians University, Munich,

31   Germany

32   12 -Department of Psychiatry, University of Münster, Münster, Germany

33   13 -Melbourne Neuropsychiatry Centre, University of Melbourne, Melbourne, Australia

34   14 -Department of Psychiatry and Psychotherapy, Faculty of Medicine and University

35   Hospital, University of Cologne, Cologne, Germany

36   15 -Department of Psychiatry, University of Turku, Turku, Finland

37   16 -Department of Psychiatry and Psychotherapy, Heinrich-Heine University Düsseldorf,

38   Düsseldorf, Germany

39   17 -IRCCS "E. Medea" Scientific Institute, San Vito al Tagliamento (Pn), Italy

40   18 -Early Intervention Service, Birmingham Women's and Children's NHS Foundation trust,

41   United Kingdom

Lalousis et al.   Transdiagnostic features, comorbidity and classification

54   **Corresponding Author**:

55   Mr. Paris Alexandros Lalousis, BSc, MSc

56   School of Psychology

57   University of Birmingham, 52 Pritchatts Road

58   B15 2SA, Birmingham, United Kingdom

59   pal532@student.bham.ac.uk

60

61   **Word Count Manuscript**

62   Abstract: 250

63   Introduction: 635

64   Methods: 877

65   Results: 750

66   Discussion: 1264

67   Figure Legends: 223

68   Total: 3999 (including abstract, text body, and figure and table legends)

69

70   **Supplementary Methods and Results**: 1 file

71

72

73

74

75

76

77

78

79

80

81

82

83

84

Lalousis et al.   Transdiagnostic features, comorbidity and classification

85   **Abstract**

86   Diagnostic heterogeneity within and across psychotic and affective disorders challenges

87   accurate treatment selection, particularly in early stages. Delineation of shared and distinct

88   illness features at the phenotypic and brain levels may inform the development of more

89   precise differential diagnostic tools. We aimed to identify prototypes of depression and

90   psychosis to investigate their heterogeneity, with common, comorbid transdiagnostic

91   symptoms. Analysing clinical/neurocognitive and grey matter volume (GMV) data from the

92   PRONIA database, we generated prototypic models of recent-onset depression (ROD) vs.

93   recent-onset psychosis (ROP) by training support-vector machines to separate patients with

94   ROD from patients with ROP, who were selected for absent comorbid features (pure groups).

95   Then, models were applied to patients with comorbidity, i.e., ROP with depressive symptoms

96   (ROP+D) and ROD participants with sub-threshold psychosis-like features (ROD+P), to

97   measure their positions within the affective-psychotic continuum. All models were

98   independently validated in a replication sample. Comorbid patients were positioned between

99   pure groups, with ROP+D patients being more frequently classified as ROD compared to

100 pure ROP patients (clinical/neurocognitive model: χ^2=14.874; p<0.001; GMV model:

101 χ^2=4.933; p=0.026). ROD+P patient classification did not differ from ROD

102 (clinical/neurocognitive model: χ^2=1.956; p=0.162; GMV model: χ^2=0.005; p=0.943).

103 Clinical/neurocognitive and neuroanatomical models demonstrated separability of prototypic

104 depression from psychosis. The shift of comorbid patients towards the depression prototype,

105 observed at the clinical and biological levels, suggests that psychosis with affective

106 comorbidity aligns more strongly to depressive rather than psychotic disease processes.

107 Future studies should assess how these quantitative measures of comorbidity predict

108 outcomes and individual responses to stratified therapeutic interventions.

109

110

111

112

113

114

Lalousis et al.   Transdiagnostic features, comorbidity and classification

115   **Introduction:**

116   Treatments for mental illness are currently based on categorical structures built on patterns of

117   syndromes and their course, rather than aetiology[1]. The biological and clinical overlaps

118   between these syndromes, and significant heterogeneity in outcomes, has become more

119   apparent in recent years[2–4]. Advancement in both pharmacological and psychotherapeutic

120   interventions has stalled, potentially as a result of continued focus on invalid disease

121   categories[5,6]. The need for better treatments is particularly acute in psychosis and depression,

122   which constitute major mental health challenges to the world's population[7–13]. The legacy of

123   a Jaspers based hierarchical approach to symptom structures suggests that positive psychotic

124   symptoms are of primary importance within psychotic spectrum disorders[14]. Yet the

125   categorical and hierarchical division of psychotic disorders into affective and non-affective

126   has been contested for decades[15], with clear demonstration of the presence of affective

127   symptoms in psychosis and psychotic symptoms in affective disorders[8,11,16].

128   Heterogeneity is particularly noticeable in early and developing stages of illness, with high

129   prevalence of affective symptoms across disorders[17]. The comorbidity of depression in early

130   psychosis has been largely regarded as secondary to the primary disorder (psychosis),

131   reinforcing a categorical, hierarchical approach[18,19]. There are increasing calls to use

132   empirical evidence to develop alternative aetiologically informed structures[20]. The use of

133   multidimensional item response modelling to predict psychosis biotypes has been shown to

134   transcend traditional diagnostic boundaries; with suggestion of an underlying transdiagnostic

135   dimension across psychotic diagnoses[21–23]. However, there are valid reasons why a

136   categorical approach to mental illness has persisted; a significant number of individuals'

137   presentation will 'fit' within distinct categories of mental illness and the course and outcome

138   of their treatment can be predicted from such diagnostic structures[24]. When disorders are fully

139   formed, course as well as symptom structure enable clearer distinction between categories.

140   Imaging studies also show shared areas of interest, including the hippocampus and

141   cerebellum[32], the prefrontal cortex and insula[33] and both depression and psychosis have been

142   associated with heightened brain activation in regions central to emotional processing[33–35]

143   with similarities particularly prominent in the early stages of illness[25,33–35].

144   The distinction between depression and schizophrenia is possible by structural brain data, but

145   also, more challenging in early stages of illness when symptoms and course are more

146   heterogeneous[25]. Transdiagnostic processes of mental health disorders are descriptively

Lalousis et al.   Transdiagnostic features, comorbidity and classification

147 transdiagnostic (i.e. being present in multiple disorders, without regard to how or why) or

148 mechanistically transdiagnostic (i.e. reflecting neurobiological, physiological, or functional

149 mechanisms)[26,27]. Both depression and psychosis are associated with transdiagnostic features

150 of working memory, executive functioning, and verbal fluency deficits[28]. [30]. The importance

151 of certain mechanistically transdiagnostic symptoms is potentially hidden in categorical

152 structures, and they remain under-investigated[31].

153 Complex psychopathology and heterogeneity in developing mental health disorders presents

154 the opportunity of fuller exploration of the significance of potential transdiagnostic

155 symptoms, to provide further insight into aetiopathogenetic pathways of symptoms and

156 through this to advance diagnostic structures[36,37]. However, novel approaches and powerful

157 statistical tools such as machine learning techniques could help provide this deeper

158 understanding by detecting complex patterns of data across diagnostic structures, and the

159 delineation of shared and distinct features of these illnesses at the phenotypic and brain

160 levels. This may inform the development of more precise differential diagnostic tools and

161 improve the development of new treatments[36].

162 This study aimed to identify prototypes of pure depression and pure psychosis in order to

163 investigate the heterogeneity of depression and psychosis with common, comorbid

164 transdiagnostic symptoms. We hypothesized that developed models would correctly classify

165 diagnostic groups without comorbid symptoms, in keeping with evidence of the utility in

166 categorical diagnostic structure, and that grey matter volume (GMV) would add classification

167 accuracy. We further hypothesised that a reduction of classification accuracy would be seen

168 in groups with comorbid symptoms. Exploration and the delineation of shared and distinct

169 features at both the phenotypic and biological levels may potentially inform future

170 development of more precise treatments.

171 **Materials and Methods**

172 **Study design**

173 Data were taken from the discovery and replication samples of the PRONIA study, an EU-

174 FP7 funded seven centre study aiming to optimize candidate biomarkers for the prediction

175 and staging of mental health disorders. Details of the PRONIA study sites, recruitment

176 protocol and quality control procedures are described in a previous publication[40] and in the

177 supplementary methods (1.1, 1.2, 1.3, tables S2, S3).

Lalousis et al.   Transdiagnostic features, comorbidity and classification

178   **Inclusion and Exclusion Criteria**

179   The general inclusion criteria for the study were: (1) age between 15 and 40 years, (2)

180   sufficient language skills for participation, (3) capacity to provide informed consent/assent.

181   General exclusion criteria were: (1) an IQ below 70, (2) current or past head trauma with loss

182   of consciousness (> 5 minutes), (3) current or past known neurological or somatic disorders

183   potentially affecting structure or functioning of the brain, (4) current or past alcohol

184   dependence, (5) polysubstance dependence within the past six months, and (6) any medical

185   indication against MRI. ROP and ROD inclusion criteria can be found in the supplement

186   (1.1).

187   **Group identification**

188   Pure ROP: any ROP patient who had a Beck Depression Inventory-II (BDI-II) score of 13 or

189   lower, which is indicative of absent or minimal depressive symptoms[41,42]. Pure ROD: any

190   ROD patient who had a Positive and Negative Symptom Scale (PANSS)[43] positive subscale

191   score of no more than 7 and no Structured Interview of Psychosis-risk Syndromes positive

192   (SIPS-P) severity score of 3 or more on any item.

193   ROP with depressive symptoms (ROP+D): any ROP participant with a BDI-II score of 14 or

194   more. ROD with psychotic symptoms (ROD+P): any ROD participant with a SIPS positive

195   item score of 2 or more and a Schizophrenia Proneness Instrument-Adult version (SPI-A)

196   Cognitive Disturbances (COGDIS) item score of 3 or more.

197   Twenty-four ROP and 31 ROD patients from the discovery sample and 21 ROP and 53 ROD

198   patients were not included in the analysis due to not meeting group identification criteria or

199   not having neuroimaging data.

200   **MRI imaging data acquisition, quality control, and preprocessing**

201   Participants underwent a multi-modal MRI protocol. A minimal harmonization protocol,

202   which the MR sequences across the different scanners had to comply with is described in the

203   supplementary methods (1.3). In the current study, T1-weighted structural MRI (sMRI)

204   images of the participants were analyzed. The sMRI images of six healthy travelling

205   volunteers who were scanned at all sites with same parameters were also analysed as part of a

206   calibration study. The images were processed using the open source CAT12 toolbox (version

207   r1155; http://dbm.neuro.uni.jena.de/cat12/) (see supplementary methods 1.4). Employing

Lalousis et al.   Transdiagnostic features, comorbidity and classification

208   generalization theory[44,45], a between-site voxel reliability map (G coefficient map) was

209   computed from the analysis of the GMV maps that were derived from the calibration study.

210   During our neuroimaging based machine learning analyses the G coefficient maps were used

211   for reliability-based voxel masking.

Classification Models

213   Using the pure ROP and ROD groups, a Support Vector Machine (SVM) classification model

214   was built using individual item scores from broad clinical and neurocognitive tests that assess

215   features commonly occurring in psychosis and depression including anhedonia, social

216   functioning and cognition deficits (see supplementary methods 1.8). In total, 151 features

217   were included in the model. The trained pure classification model was then applied to the

218   comorbid ROP+D and ROD+P groups to determine the classification accuracy to their

219   primary diagnosis.

220   A second model using GMV whole-brain voxel-wise data as features was developed in the

221   pure ROD and ROP groups and then applied to the ROP+D and ROD+ P groups. The

222   developed clinical/cognitive and GMV models were combined by using decision values from

223   the pure clinical and pure GMV models (in order to build a model that learns from the meta-

224   data) in a stacking-based data fusion framework[46,47]. Finally, all the models were applied to

225   an independent replication sample.

Support Vector Machine Learning Analysis:

227   The machine learning analysis of pre-processed data (see supplementary methods 1.5) was

228   performed using NeuroMiner (version 1.0; https://github.com/neurominer-git). A repeated

229   nested pooled cross-validation (CV) was used with 10 outer CV2 permutations, 10 outer CV2

230   folds, 10 inner CV1 permutations, and 10 inner CV1 folds.

231   Imbalanced learning was corrected for, by increasing the C value in the minority class by

232   multi-plying it by the inverse ratio of the training class sizes. A linear kernel was used with

233   eleven C values (0.0156, 0.0312, 0.0625, 0.1250, 0.2500, 0.5000, 1, 2, 4, 8, and 16) in order

234   to optimize the choice of C value and create an ensemble of predictive models to be applied

235   to the CV_2 data to produce a single average robust prediction.

236   Balanced accuracy (BAC) regularized by SVM model complexity was used as a criterion for

237   the hyperparameter optimization (see supplementary methods 1.6).

Lalousis et al.   Transdiagnostic features, comorbidity and classification

238 Finally, a stacking-based data fusion framework[46–49] was used to examine whether the

239 combination of the clinical/neurocognitive and the neuroimaging-based models would

240 provide a superior classification accuracy (see supplementary methods 1.7).

Independent Validation

242 All our models were validated in our independent replication sample (N=262) (see

243 supplementary methods table S4) which was collected at a different timescale. The same

244 group identification criteria were applied.

Supplementary analyses

246 A number of supplementary exploratory analyses, including correlation analyses between

247 decision scores from our models, association and comparison analyses between correctly and

248 mis-classified patients, GMV comparison between groups, decision score group comparisons,

249 and regression analyses with 9 month functional outcomes were conducted and can be found

250 in the supplement (Section 2).

Results:

Demographic Information

253 Data from 154 participants with ROP and 146 patients with ROD were included in the

254 analysis as our Discovery sample. Thirty-eight ROP patients were included in the pure ROP

255 group and 90 ROD patients in the pure ROD group. The mean age of the pure ROP group

256 was 26.5 [SD 6.8]) and the mean age of the pure ROD group was 26.5 [SD 6.6]). There were

257 25 male and 13 female patients in the pure ROP group and 45 male and 45 female patients in

258 the pure ROD group.

259 Ninety-two ROP subjects were included in the ROP+D group and 25 ROD subjects in the

260 ROD+P group. The mean age of the ROP+D group was 26.5 [SD 6.8] and the mean age of

261 the ROD+P group was 23.8 [SD 3.9]. There were 57 males and 35 females in the ROP+D

262 group and 12 males and 13 females in the ROD+P group. A summary of demographic

263 information is provided in table 1.

264 The independent validation sample consisted of 161 patients with ROP and 131 patients with

265 ROD. Fifty ROP patients were included in the pure ROP group and 53 ROD patients in the

Lalousis et al.   Transdiagnostic features, comorbidity and classification

266   pure ROD group. The ROP+D and ROD+P groups consisted of 90 and 25 patients

267   respectively. A full description is provided in the supplement (Table S4).

Machine Learning Analyses

Internal Validation of the Pure Group Differential Classifiers

Clinical and Neurocognitive Data

271   A repeated nested pooled cross validation model with classifiers of clinical and cognitive

272   variables predicted pure diagnostic groups with a balanced accuracy (BAC) of 79.3%; 95%

273   CI [77.2, 82.3] and an area under the curve (AUC) of 0.86 (Table 2 and Figure 1a).

274   Assignment to the ROP category by the clinical classifier was driven by reduced scores in the

275   RSA and elevated scores in the WSS, RSA, and SPIA. ROD group classification was

276   informed by increased scores in the SPIA, WSS, together with reduced scores in the DSST.

277   The contribution of the features was calculated by feature weights and by cross-validation

278   ratio (Figure 2).

GMV Data

280   The repeated nested pooled cross-validation model using sMRI to predict diagnostic group in

281   pure ROP and ROD produced a BAC of 62.5%; 95% CI [58.8, 64.0] and an AUC of 0.70

282   (Table 2 and Figure 1b). ROP patients showed pronounced reductions in the thalamus and the

283   cerebellum, whereas depressed patients showed orbitofrontal, limbic and paralimbic volume

284   reductions (Figure 3).

Stacking

286   Combining the outputs of the clinical predictors and sMRI using stacked generalization

287   predicted diagnostic group with a BAC of 79.5%; 95% CI [77., 81.9] and an area under the

288   curve (AUC) of 0.87 (Table 2 and Figure 1c).

Separability of Comorbid Groups

290   **Clinical/Neurocognitive Data**: The trained pure classification system comprising the

291   collection of 11 clinical/neurocognitive models generated by the repeated nested cross-

292   validation scheme on pure groups was then applied to the comorbid groups (ROP+D and

293   ROD+P) to produce decision scores measuring ROP vs. ROD likeness. This model had a

294   BAC of 62.5% and an AUC of 0.66. Misclassifications showed a directionality toward the

Lalousis et al.   Transdiagnostic features, comorbidity and classification

295   ROD group, with 63% of ROP+D patients being classified as ROD; $Z=1.276$, $p=.0385$) (see

296   Table 2 and Figure 1a). ROP+D patients were more frequently classified as ROD compared

297   to pure ROP patients ($\chi^2=14.874$; $p<0.001$). In contrast, the assignment precision of ROD+P

298   and ROD patients did not differ ($\chi^2=1.956$; $p=0.162$).

GMV Data

300   The trained pure classification system (comprising of 11 GMV models generated by the

301   repeated nested cross-validation scheme on pure groups) was then applied to the comorbid

302   groups (ROP+D and ROD+P) to produce decision scores measuring ROP vs. ROD likeness.

303   This produced a BAC of 47.8% and AUC of 0.43. Misclassifications showed a directionality

304   toward the ROD group, with 80.4% of ROP+D patients being classified as ROD; $Z=.713$,

305   $p=.344$) (see Table 2 and Figure 1b). Similarly to the clinical/neurocognitive model ROP+D

306   patients were more frequently classified as ROD compared to pure ROP patients ($\chi^2=4.933$;

307   $p=0.026$). In contrast, the assignment precision of ROD+P and ROD patients did not differ

308   ($\chi^2=0.005$; $p=0.943$).

Stacking

310   When applied to the comorbid groups, the combined model predicted diagnostic group with a

311   BAC of 58.5% and an area under the curve (AUC) of 0.66 (see Table 1 and Figure 1c).

Independent Validation

313   Application of our models to the independent validation sample replicated findings very well

314   (pure clinical and neurocognitive model BAC 76.2; pure imaging model BAC 49.9; pure

315   stacking model BAC 78.2). Full results from the independent validation analysis can be

316   found in the supplement (Table S5).

Supplementary Analyses

318   See supplement S2 for additional exploratory analyses results.

Discussion:

320   Using repeated nested cross-validation techniques we built classification models based on

321   transdiagnostic clinical and neurocognitive features and GMV data, together with a combined

322   model integrating all data modalities, to classify prototype diagnostic groups of ROD and

323   ROP participants without comorbidity. Eighty-seven per cent of patients with pure ROP and

Lalousis et al.   Transdiagnostic features, comorbidity and classification

324 ROD were accurately ascribed to their diagnostic group. Applying this model to groups with

325 comorbidity, 88% of patients with ROD and psychotic features were ascribed to their primary

326 diagnostic group (depression) whereas only 37% of patients with ROP and depressive

327 features were ascribed to their primary diagnostic group (psychosis). The shift of comorbid

328 psychosis patients towards the depression prototype was observed both at the clinical and

329 biological levels. This suggests that when comorbid with affective symptoms, psychoses

330 align more strongly to the disease processes of depressive than psychotic disorders. Using

331 GMV measures only for classification, comorbid groups ROP patients with depressive

332 symptoms largely resembled pure ROD. Results were generalisable to our independent

333 validation sample.

334 Our findings suggest that clinical and neurocognitive transdiagnostic symptoms may have

335 differential weight within psychosis and depression presentation with and without

336 comorbidity. In pure groups, these symptoms could be seen as under the hierarchical

337 umbrella of psychosis or depression, and may accurately reflect underlying pathology in

338 these groups. However, in the face of comorbidity, transdiagnostic symptoms lean more to

339 the depression domain. Given the prevalence and importance of depressive comorbidity in

340 early psychosis this may support a model where depression could be more intrinsically

341 important than is currently considered in early phases of illness[4,29]. When participants with

342 ROP exhibited even mild depressive symptoms, their GMV classification was more likely to

343 lie within the depression group than psychosis suggesting a potential depressive biological

344 phenotype that exists in the psychosis spectrum. The implications of these findings suggest

345 that understanding heterogeneity of brain structures may need to include specific focus on

346 symptoms that may often be masked by more acute (e.g. positive) symptoms and simple

347 solutions that some symptoms may be transdiagnostic, potentially belie the complexity of

348 individual aetiology and psychopathology. Furthermore, these findings indicate a need to

349 rethink current diagnostic classification to better reflect the biological reality and eventually

350 develop better treatment options.

351 A classical diagnostic hierarchy in the structure of personal illness, reflected in current

352 nosological classification systems, posits that mental disorders of the primary diagnosis have

353 more weight and primacy over symptoms from lower classes, which are seen as secondary or

354 comorbid[18,19]. We found that comorbid symptoms affect diagnostic structures in different

355 ways. Sub-clinical psychotic-like symptoms appeared to not alter the signature of ROD

Lalousis et al.   Transdiagnostic features, comorbidity and classification

356   patients whilst depressive symptoms had a profound effect on ROP patients' classification

357   accuracy.

358   Our findings add evidence to the debate around the validity of the system on which the DSM

359   is built upon[50], suggesting that a descriptivist position in the diagnosis of mental disorder is

360   not sufficient and that a novel multivariate approach of mental disorder is more appropriate.

361   Comorbidity in psychiatric disorders presents a clinical as well as nosological challenge.

362   There may be significant interplay between sets or clusters of symptoms over the

363   development of disorder from prodrome via onset to potential chronicity. The frequency of

364   depression within ROP may be a primary driver, rather than being a secondary symptom. If

365   identified correctly, novel symptoms may be new targets that if treated effectively, ameliorate

366   the other, e.g. positive symptoms.

367   Recent onset disorders may constitute groups of phenotypically highly individual symptoms

368   with underlying aetiopathology, and it may be that personalized treatments could be tailored

369   accordingly. Our SVM classification model found that different types of anhedonia (social

370   and physical) were important in the classification of both psychosis and depression.

371   Anhedonia has been suggested as a possible biomarker for depression[51] and has been found to

372   be associated with decreased activation in ventral basal ganglia areas, the dorsal anterior

373   cingulate, middle frontal gyrus, and medial frontal gyrus both in schizophrenia and

374   depression[52]. Our GMV model revealed that orbitofrontal areas were higher weighted in the

375   classification to the ROD group.

376   The relationship between psychotic and affective symptoms has been central to the dilemma

377   of psychiatric classification. Substantial clinical and genomic evidence shows that

378   schizophrenia and affective disorders may be distributed across a dimensional spectrum[53].

379   However, the dimensional spectra model does not allow for either the clinical reality of a

380   complex and changing symptom profile, nor the investigation of clinical features commonly

381   seen across all disorders; some of which may be of primary importance. Concerning the

382   neurobiology of schizophrenia and depression, the majority of previous studies are based in

383   subjects with depression or psychosis, but only rarely in both[25,54] and not previously in highly

384   mixed recent onset comorbid disorders. Our results suggest that while GMV showed some

385   distinction in prototype (pure) groups, when presented with complex comorbid groups, which

386   may be the majority in clinical practice, there was a significant lack of any point of rarity

387   between disorders. This builds on previous work suggesting distinction is more challenging,

Lalousis et al.   Transdiagnostic features, comorbidity and classification

388   but also that neuroimaging based data do not support categorical classification in recent onset

389   disorders [25]. In recent onset disorders, early cognitive processes related to depression can

390   both drive other more severe symptoms and/or be seen in isolation: for example, anhedonia

391   could be an early indication of negative symptom clusters or a core feature of co-morbid

392   depression[55].

393

394   **Strengths and Limitations**

395   The strengths of the present analysis include sufficiently large data, robust collection of

396   clinical and imaging data from both depression and psychosis groups, independent validation

397   analysis together with a novel approach to a challenging and essential clinically relevant

398   research question, which speaks to the validity of diagnoses as the cornerstone of psychiatric

399   practice. Our results however should be interpreted with certain caution due to limitations

400   with the study. Regarding our definition of the ROD+P group we used a SIPS-P item score of

401   2 or more which is not a marker of formal psychotic symptoms, and thus would only measure

402   low levels of psychotic-like symptoms. However to supplement this we used a SPI-A

403   COGDIS item score of 3 or more. We did not include core symptomatology measures such as

404   the PANSS and the BDI in our features due to the fact that primary groups are defined with

405   these measures, and therefore including them would risk a circular analysis. Finally, there

406   were more subjects identified in the pure ROD and comorbid ROP groups; ideally groups of

407   equal size would have been used. Nevertheless, we addressed this imbalance in our analysis,

408   by increasing the C value in the minority class by multiplying it by the inverse ratio of the

409   training class sizes.

410   **Conclusions**

411   Findings from this large, multi-modal, replicated machine learning classification study in

412   recent onset disorders suggest that whilst there may be a small subset of prototypically pure

413   individuals with clear categorical disorder, the majority of patients share a number of

414   transdiagnostic features, primarily from the depression domain. Brain structure of psychosis

415   patients with co-morbid depressive symptoms largely resembles that of depression. The

416   increasing interest in heterogeneity of early disorders and transdiagnostic symptoms as novel

417   treatment targets needs to be fully informed of potential depression related co-morbidity. Our

418   analysis in recent onset groups also highlight that both categorical and transdiagnostic

419   approaches may ultimately fail at an individual patient level, as neither recognise the

Lalousis et al.   Transdiagnostic features, comorbidity and classification

420   possibility of pluripotent pathways that are both stage and context dependent. The

421   implications, particularly for early intervention and prevention in mental health disorders is

422   that ultimately a personalised medicine approach, encompassing the full potential of

423   comorbidity, may be necessary to improve outcomes. Future studies should investigate the

424   utility of targeting such transdiagnostic depression features to elucidate their prognostic

425   value, and develop new treatments.

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

Lalousis et al.   Transdiagnostic features, comorbidity and classification

444   **References**

445   1.   Bzdok D, Meyer-Lindenberg A. Machine Learning for Precision Psychiatry: Opportunities and
446   Challenges. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2018;3(3):223-230.
447   doi:10.1016/j.bpsc.2017.11.007

448   2.   Kambeitz J, Kambeitz-Ilankovic L, Leucht S, et al. Detecting Neuroimaging Biomarkers for
449   Schizophrenia: A Meta-Analysis of Multivariate Pattern Recognition Studies.
450   *Neuropsychopharmacology*. 2015;40(7):1742-1751. doi:10.1038/npp.2015.22

451   3.   Kambeitz J, Cabral C, Sacchet MD, et al. Detecting Neuroimaging Biomarkers for Depression: A
452   Meta-analysis of Multivariate Pattern Recognition Studies. *Biol Psychiatry*. 2017;82(5):330-338.
453   doi:10.1016/j.biopsych.2016.10.028

454   4.   Upthegrove R, Lalousis P, Mallikarjun P, et al. The Psychopathology and Neuroanatomical
455   Markers of Depression in Early Psychosis. *Schizophr Bull*. Published online July 7, 2020:sbaa094.
456   doi:10.1093/schbul/sbaa094

457   5.   Wong EHF, Yocca F, Smith MA, Lee C-M. Challenges and opportunities for drug discovery in
458   psychiatric disorders: The drug hunters' perspective. *Int J Neuropsychopharmacol*.
459   2010;13(9):1269-1284. doi:10.1017/S1461145710000866

460   6.   Hofmann SG, Asnaani A, Vonk IJJ, Sawyer AT, Fang A. The Efficacy of Cognitive Behavioral
461   Therapy: A Review of Meta-analyses. *Cogn Ther Res*. 2012;36(5):427-440. doi:10.1007/s10608-
462   012-9476-1

463   7.   Charlson FJ, Ferrari AJ, Santomauro DF, et al. Global Epidemiology and Burden of
464   Schizophrenia: Findings From the Global Burden of Disease Study 2016. *Schizophr Bull*.
465   2018;44(6):1195-1203. doi:10.1093/schbul/sby058

466   8.   WHO | Depression and Other Common Mental Disorders. WHO. Published 2017. Accessed
467   May 20, 2019.
468   http://www.who.int/mental_health/management/depression/prevalence_global_health_esti
469   mates/en/

470   9.   Chong HY, Teoh SL, Wu DB-C, Kotirum S, Chiou C-F, Chaiyakunapruk N. Global economic
471   burden of schizophrenia: a systematic review. *Neuropsychiatr Dis Treat*. 2016;12:357-373.
472   doi:10.2147/NDT.S96649

473   10.   Olesen J, Gustavsson A, Svensson M, Wittchen H-U, Jönsson B. The economic cost of brain
474   disorders in Europe. *Eur J Neurol*. 2012;19(1):155-162. doi:10.1111/j.1468-1331.2011.03590.x

475   11.   Conley RR, Ascher-Svanum H, Zhu B, Faries D, Kinon BJ. The Burden of Depressive Symptoms in
476   the Long-Term Treatment of Patients With Schizophrenia. *Schizophr Res*. 2007;90(1-3):186-
477   197. doi:10.1016/j.schres.2006.09.027

478   12.   Sobocki P, Jönsson B, Angst J, Rehnberg C. Cost of Depression in Europe. *J Ment Health Policy
479   Econ*. 2006;9(2):87-98.

480   13.   Dieleman JL, Baral R, Birger M, et al. US Spending on Personal Health Care and Public Health,
481   1996-2013. *JAMA*. 2016;316(24):2627-2646. doi:10.1001/jama.2016.16885

Lalousis et al.   Transdiagnostic features, comorbidity and classification

482   14.   Bürgy M. The Concept of Psychosis: Historical and Phenomenological Aspects. *Schizophr Bull*.
483   2008;34(6):1200-1210. doi:10.1093/schbul/sbm136

484   15.   Quattrone D, Di Forti M, Gayer-Anderson C, et al. Transdiagnostic dimensions of
485   psychopathology at first episode psychosis: findings from the multinational EU-GEI study.
486   *Psychol Med*. 2019;49(08):1378-1391. doi:10.1017/S0033291718002131

487   16.   Ohayon MM, Schatzberg AF. Prevalence of Depressive Episodes With Psychotic Features in the
488   General Population. *Am J Psychiatry*. 2002;159(11):1855-1861.
489   doi:10.1176/appi.ajp.159.11.1855

490   17.   Rickwood D, Van Dyke N, Telford N. Innovation in youth mental health services in Australia:
491   common characteristics across the first headspace centres. *Early Interv Psychiatry*.
492   2015;9(1):29-37. doi:10.1111/eip.12071

493   18.   Birchwood M, Iqbal Z, Upthegrove R. Psychological pathways to depression in schizophrenia:
494   studies in acute psychosis, post psychotic depression and auditory hallucinations. *Eur Arch*
495   *Psychiatry Clin Neurosci*. 2005;255(3):202-212. doi:10.1007/s00406-005-0588-4

496   19.   Foulds GA, Bedford A. Hierarchy of classes of personal illness. *Psychol Med*. 1975;5(2):181-192.
497   doi:10.1017/S0033291700056452

498   20.   Clark LA, Cuthbert B, Lewis-Fernández R, Narrow WE, Reed GM. Three Approaches to
499   Understanding and Classifying Mental Disorder: ICD-11, DSM-5, and the National Institute of
500   Mental Health's Research Domain Criteria (RDoC). *Psychol Sci Public Interest*. 2017;18(2):72-
501   145. doi:10.1177/1529100617727266

502   21.   Maj M. The need for a conceptual framework in psychiatry acknowledging complexity while
503   avoiding defeatism. *World Psychiatry Off J World Psychiatr Assoc WPA*. 2016;15(1):1-2.
504   doi:10.1002/wps.20291

505   22.   Tamminga CA, Ivleva EI, Keshavan MS, et al. Clinical phenotypes of psychosis in the Bipolar-
506   Schizophrenia Network on Intermediate Phenotypes (B-SNIP). *Am J Psychiatry*.
507   2013;170(11):1263-1274. doi:10.1176/appi.ajp.2013.12101339

508   23.   Dwyer DB, Kalman JL, Budde M, et al. An Investigation of Psychosis Subgroups With Prognostic
509   Validation and Exploration of Genetic Underpinnings: The PsyCourse Study. *JAMA Psychiatry*.
510   2020;77(5):523-533. doi:10.1001/jamapsychiatry.2019.4910

511   24.   Potuzak M, Ravichandran C, Lewandowski KE, Ongür D, Cohen BM. Categorical vs dimensional
512   classifications of psychotic disorders. *Compr Psychiatry*. 2012;53(8):1118-1129.
513   doi:10.1016/j.comppsych.2012.04.010

514   25.   Koutsouleris N, Meisenzahl EM, Borgwardt S, et al. Individualized differential diagnosis of
515   schizophrenia and mood disorders using neuroanatomical biomarkers. *Brain*.
516   2015;138(7):2059-2073. doi:10.1093/brain/awv111

517   26.   Harvey AG, Murray G, Chandler RA, Soehner A. Sleep disturbance as transdiagnostic:
518   Consideration of neurobiological mechanisms. *Clin Psychol Rev*. 2011;31(2):225-235.
519   doi:10.1016/j.cpr.2010.04.003

520   27.   Harvey AG, Watkins E, Mansell W. *Cognitive Behavioural Processes Across Psychological*
521   *Disorders: A Transdiagnostic Approach to Research and Treatment*. Oxford University Press;
522   2004.

523   28.   Moritz S, Birkner C, Kloss M, et al. Executive functioning in obsessive–compulsive disorder,
524   unipolar depression, and schizophrenia. *Arch Clin Neuropsychol*. Published online 2002:7.

525   29.   Upthegrove R, Marwaha S, Birchwood M. Depression and Schizophrenia: Cause, Consequence,
526   or Trans-diagnostic Issue? *Schizophr Bull*. 2017;43(2):240-244. doi:10.1093/schbul/sbw097

527   30.   Carrà G, Crocamo C, Bartoli F, et al. The mediating role of depression in pathways linking
528   positive and negative symptoms in schizophrenia. A longitudinal analysis using latent variable
529   structural equation modelling. *Psychol Med*. Published online March 8, 2019:1-9.
530   doi:10.1017/S0033291719000321

531   31.   Krueger RF, Eaton NR. Transdiagnostic factors of mental disorders. *World Psychiatry*.
532   2015;14(1):27-29. doi:10.1002/wps.20175

533   32.   Meisenzahl EM, Seifert D, Bottlender R, et al. Differences in hippocampal volume between
534   major depression and schizophrenia: a comparative neuroimaging study. *Eur Arch Psychiatry*
535   *Clin Neurosci*. 2010;260(2):127-137. doi:10.1007/s00406-009-0023-3

536   33.   Busatto GF. Structural and functional neuroimaging studies in major depressive disorder with
537   psychotic features: a critical review. *Schizophr Bull*. 2013;39(4):776-786.
538   doi:10.1093/schbul/sbt054

539   34.   Kumari V, Peters E, Guinn A, et al. Mapping Depression in Schizophrenia: A Functional
540   Magnetic Resonance Imaging Study. *Schizophr Bull*. 2016;42(3):802-813.
541   doi:10.1093/schbul/sbv186

542   35.   Broome MR, He Z, Iftikhar M, Eyden J, Marwaha S. Neurobiological and behavioural studies of
543   affective instability in clinical populations: a systematic review. *Neurosci Biobehav Rev*.
544   2015;51:243-254. doi:10.1016/j.neubiorev.2015.01.021

545   36.   Dwyer DB, Falkai P, Koutsouleris N. Machine Learning Approaches for Clinical Psychology and
546   Psychiatry. *Annu Rev Clin Psychol*. 2018;14(1):91-118. doi:10.1146/annurev-clinpsy-032816-
547   045037

548   37.   Klöppel S, Abdulkadir A, Jack CR, Koutsouleris N, Mourão-Miranda J, Vemuri P. Diagnostic
549   neuroimaging across diseases. *NeuroImage*. 2012;61(2):457-463.
550   doi:10.1016/j.neuroimage.2011.11.002

551   38.   Walsh-Messinger J, Jiang H, Lee H, Rothman K, Ahn H, Malaspina D. Relative importance of
552   symptoms, cognition, and other multilevel variables for psychiatric disease classifications by
553   machine learning. *Psychiatry Res*. 2019;278:27-34. doi:10.1016/j.psychres.2019.03.048

554   39.   Jauhar S, Krishnadas R, Nour MM, Cunningham-Owens D, Johnstone EC, Lawrie SM. Is there a
555   symptomatic distinction between the affective psychoses and schizophrenia? A machine
556   learning approach. *Schizophr Res*. 2018;202:241-247. doi:10.1016/j.schres.2018.06.070

557   40.   Koutsouleris N, Kambeitz-Ilankovic L, Ruhrmann S, et al. Prediction Models of Functional
558   Outcomes for Individuals in the Clinical High-Risk State for Psychosis or With Recent-Onset

Lalousis et al.   Transdiagnostic features, comorbidity and classification

559   Depression: A Multimodal, Multisite Machine Learning Analysis. *JAMA Psychiatry*.
560   2018;75(11):1156-1172. doi:10.1001/jamapsychiatry.2018.2165

561   41.   Beck AT, Steer RA, Ball R, Ranieri W. Comparison of Beck Depression Inventories -IA and -II in
562   psychiatric outpatients. *J Pers Assess*. 1996;67(3):588-597. doi:10.1207/s15327752jpa6703_13

563   42.   Beck A, Steer R, Brown G. Manual for the Beck Depression Inventory-II (BDI-II). Published
564   online January 1, 1996. Accessed February 24, 2020.
565   https://www.scienceopen.com/document?vid=9feb932d-1f91-4ff9-9d27-da3bda716129

566   43.   Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for
567   schizophrenia. *Schizophr Bull*. 1987;13(2):261-276.

568   44.   Mushquash C, O'Connor BP. SPSS and SAS programs for generalizability theory analyses. *Behav
569   Res Methods*. 2006;38(3):542-547. doi:10.3758/BF03192810

570   45.   Robert L. Brennan. Generalizability Theory and Classical Test Theory: Applied Measurement in
571   Education: Vol 24, No 1. Published 2011. Accessed May 20, 2019.
572   https://www.tandfonline.com/doi/abs/10.1080/08957347.2011.532417

573   46.   Wolpert DH. Stacked generalization. *Neural Netw*. 1992;5(2):241-259. doi:10.1016/S0893-
574   6080(05)80023-1

575   47.   R. Polikar. Ensemble based systems in decision making - IEEE Journals & Magazine. Published
576   2006. Accessed May 20, 2019. https://ieeexplore.ieee.org/document/1688199

577   48.   Sabbagh D, Ablin P, Varoquaux G, Gramfort A, Engemann DA. Predictive regression modeling
578   with MEG/EEG: from source power to signals and cognitive states. *NeuroImage*.
579   2020;222:116893. doi:10.1016/j.neuroimage.2020.116893

580   49.   Karrer T, Ds B, B D, et al. Brain-based ranking of cognitive domains to predict schizophrenia.
581   *Hum Brain Mapp*. 2019;40(15):4487-4507. doi:10.1002/hbm.24716

582   50.   Stein DJ, Phillips KA, Bolton D, Fulford KWM, Sadler JZ, Kendler KS. What is a
583   Mental/Psychiatric Disorder? From DSM-IV to DSM-V. *Psychol Med*. 2010;40(11):1759-1765.
584   doi:10.1017/S0033291709992261

585   51.   Der-Avakian A, Markou A. The neurobiology of anhedonia and other reward-related deficits.
586   *Trends Neurosci*. 2012;35(1):68-77. doi:10.1016/j.tins.2011.11.005

587   52.   Zhang B, Lin P, Shi H, et al. Mapping anhedonia-specific dysfunction in a transdiagnostic
588   approach: an ALE meta-analysis. *Brain Imaging Behav*. 2016;10(3):920-939.
589   doi:10.1007/s11682-015-9457-6

590   53.   Smeland OB, Bahrami S, Frei O, et al. Genome-wide analysis reveals extensive genetic overlap
591   between schizophrenia, bipolar disorder, and intelligence. *Mol Psychiatry*. Published online
592   January 4, 2019. doi:10.1038/s41380-018-0332-x

593   54.   Koutsouleris N, Gaser C, Jäger M, et al. Structural correlates of psychopathological symptom
594   dimensions in schizophrenia: A voxel-based morphometric study. *NeuroImage*.
595   2008;39(4):1600-1612. doi:10.1016/j.neuroimage.2007.10.029

Lalousis et al.   Transdiagnostic features, comorbidity and classification

596   55.   Barkus E, Badcock JC. A Transdiagnostic Perspective on Social Anhedonia. *Front Psychiatry*.
597   2019;10:216. doi:10.3389/fpsyt.2019.00216

598

599   **Legends**

600   **Legend Figure 1:** Classification Performance of the Pure and Applied Clinical and

601   Neurocognitive, GMV, and Combined Models.

602   A. Pure Clinical and Neurocognitive Classification Balanced Accuracy 79.3%, Sensitivity

603   76.3%, Specificity 82.2%, AUC 0.86. Applied Clinical Classification Balanced Accuracy

604   57.6%, Sensitivity 39.1%, Specificity 76%, AUC 0.71.

605   B. Classification Performance of the Pure GMV Model and the Applied GMV Model. Pure

606   GMV Classification Balanced Accuracy 62.5%, Sensitivity 39.5%, Specificity 85.6%, AUC

607   0.70. Applied GMV Classification Balanced Accuracy 50.3%, Sensitivity 20.7%, Specificity

608   80%, AUC 0.47.

609   C. Classification Performance of the Combined Model and the Applied Combined Model.

610   Stacked Classification Balanced Accuracy 79.5%, Sensitivity 78.9%, Specificity 80%, AUC

611   0.87. Applied Stacked Classification Balanced Accuracy 64.6%, Sensitivity 53.3%,

612   Specificity 76%, AUC 0.71.

613   **Legend Figure 2:** Feature Weights and Cross-Validations Ratios of the most Significant

614   Features.

615   A. Feature Weights. Derived from 1000 random permutations of the outcome labels and

616   features.

617   B. Cross-Validation Ratio. Sum of the median weights across all CV1 folds divided by the

618   standard error.

619   **Legend Figure 3:** Significant Regions in the Imaging Classification Model. ROP GMV

620   Reductions in the Thalamus and the Cerebellum, ROD GMV Reductions in Orbitofrontal,

621   Limbic, and Paralimbic Regions.

622   **Legend Table 1:** Sample Sociodemographics. Sample Sizes, Participants per Study Site,

623   Age, Sex, Education, Partnership Status, Population Density.

Lalousis et al.   Transdiagnostic features, comorbidity and classification

624   **Legend Table 2:** Classification Performance of the Clinical and Neurocognitive, GMV, and

625   Combined Models and Validation Performance.

626

Figure 1. Classification Performance of the Pure and Applied Clinical, GMV, and Combined Models.
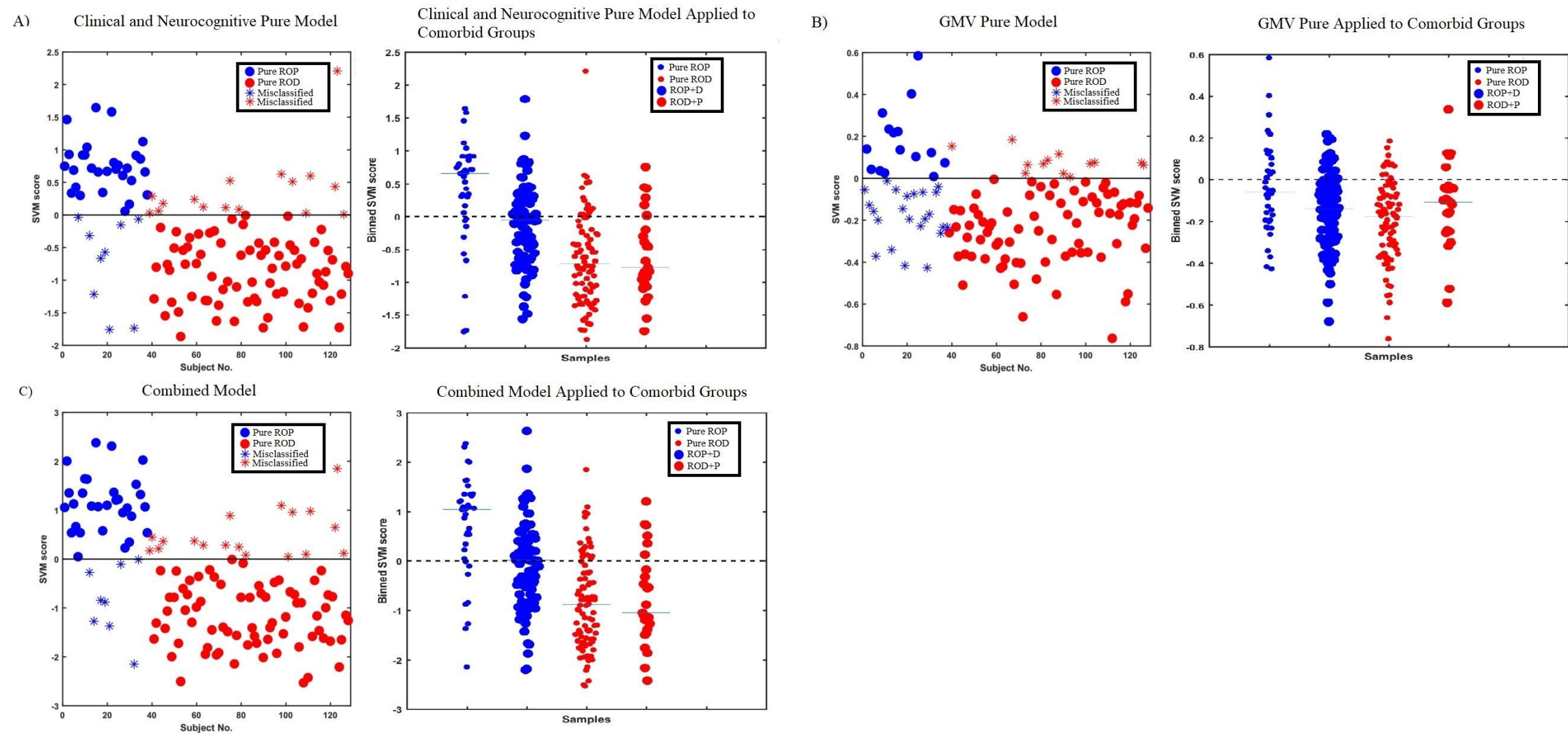
A) Pure Clinical and Neurocognitive Classification Balanced Accuracy 79.3%, Sensitivity 76.3%, Specificity 82.2%, AUC 0.86. Applied Clinical Classification Balanced Accuracy 57.6%, Sensitivity 39.1%, Specificity 76%, AUC 0.71. B) Classification Performance of the Pure GMV Model and the Applied GMV Model. Pure GMV Classification Balanced Accuracy 62.5%, Sensitivity 39.5%, Specificity 85.6%, AUC 0.70. Applied GMV Classification Balanced Accuracy 50.3%, Sensitivity 20.7%, Specificity 80%, AUC 0.47. C) Classification Performance of the Combined Model and the Applied Combined Model. Stacked Classification Balanced Accuracy 79.5%, Sensitivity 78.9%, Specificity 80%, AUC 0.87. Applied Stacked Classification Balanced Accuracy 64.6%, Sensitivity 53.3%, Specificity 76%, AUC 0.71.

Figure 2. Feature Weights and Cross-Validations Ratios of the most Significant Features.

A. Feature Weights. Derived from 1000 random permutations of the outcome labels.
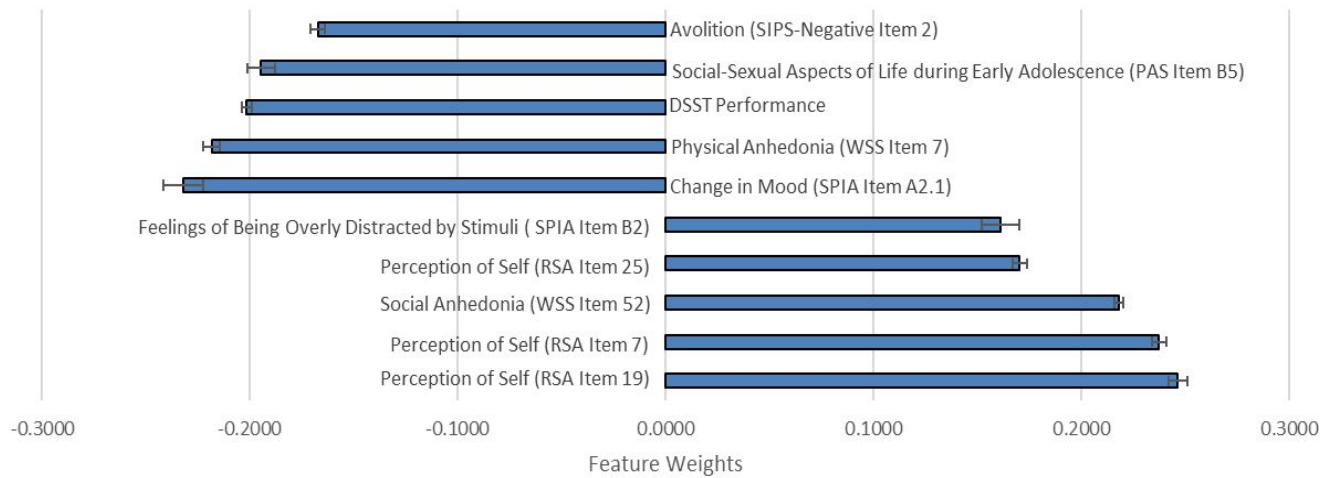
B. Cross-Validation Ratio. Sum of the median weights across all CV1 folds divided by the standard error.
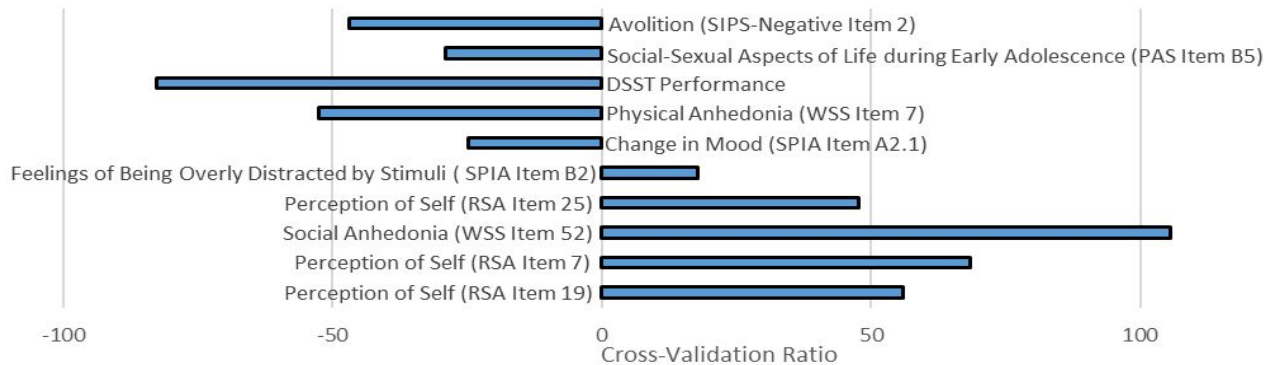
Figure 3. Significant Regions in the Imaging Classification Model. ROP GMV Reductions in the Thalamus and the Cerebellum, ROD GMV Reductions in Orbitofrontal, Limbic, and Paralimbic Regions.
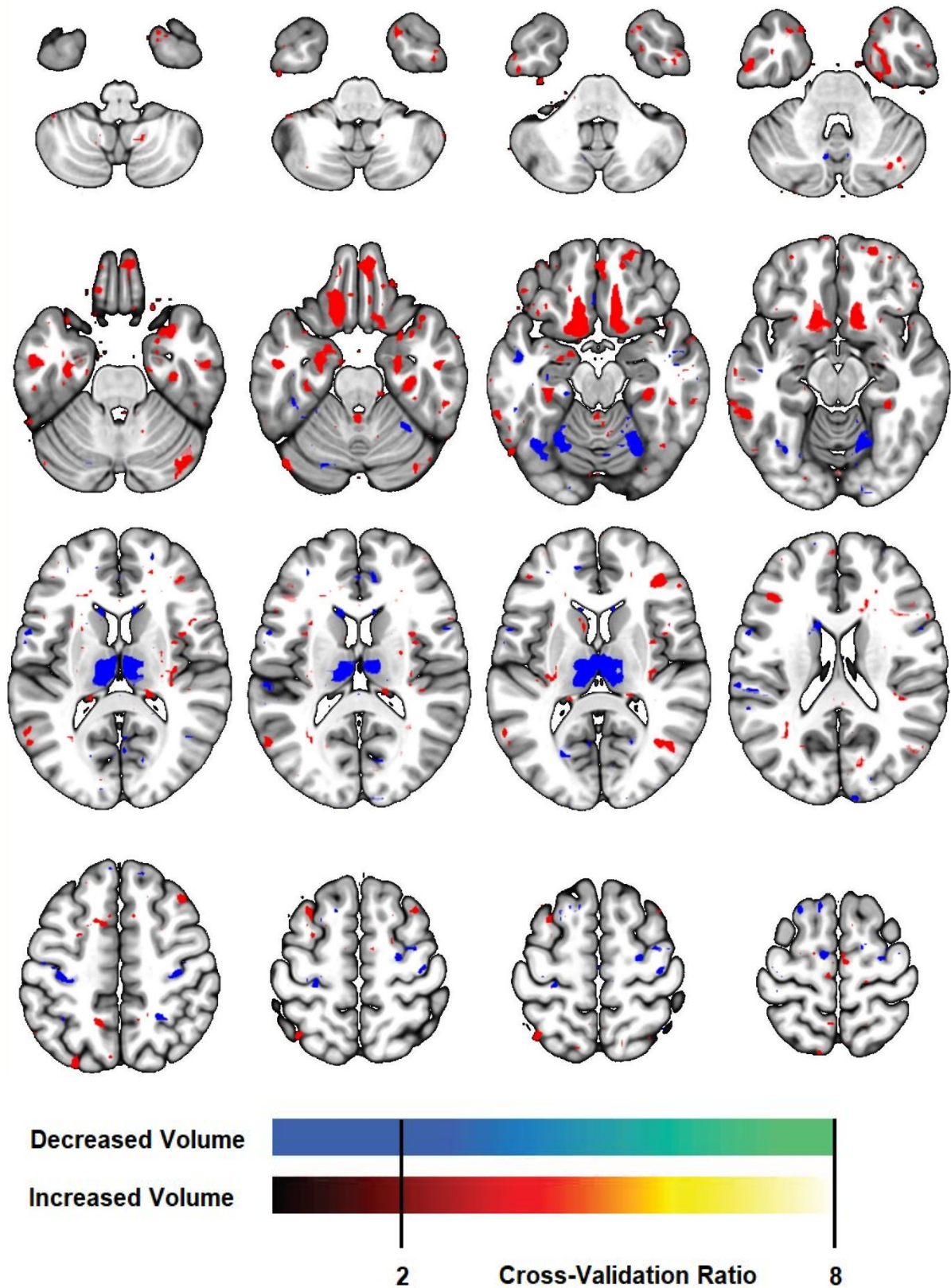
Table 1.Sample Sociodemographics. Sample Sizes, Participants per Study Site, Age, Sex, Education, Partnership Status, Population Density.

	ROP Group	ROD Group	t/χ2	P Value	Pure ROP Group	Pure ROD Group	t/z/χ2	P Value	ROP+D Group	ROD+P Group	t/z/χ2	P Value
Sample Sizes, No.	154	146			38	90			92	25		
Participants per site, No. (%)												
Basel	23 (7.7)	17 (5.7)			3 (2.3)	8 (6.9)			14 (12)	6 (5.1)		
Birmingham	10 (4.7)	10 (3.3)			4 (3.1)	0 (0)			10 (8.5)	2 (1.7)		
Cologne	27 (9)	27 (9)			4 (3.1)	19 (14.8)			22 (18.8)	3 (2.6)		
Milan	13 (4.3)	7 (2.3)	$\chi^2 = 8.9$	.257	8 (6.3)	5 (3.9)	$\chi^2 = 21.0$	.002	1 (0.9)	0 (0)	$\chi^2 = 6.1$	.517
Munich	46 (15.3)	47 (15.7)			10 (7.8)	35 (27.3)			31 (26.5)	7 (6)		
Turku	22 (7.3)	13 (4.3)			11 (8.6)	7 (5.5)			7 (6)	3 (2.6)		
Udine	12 (4)	21 (7)			2 (1.6)	12 (9.4)			3 (2.6)	3 (2.6)		
Age, Mean (SD)	24.7 (5.4)	25.5 (6.1)		.229	26.5 (6.8)	26.5 (6.6)		.959	24.6 (4.8)	23.8 (3.9)		.449
Sex (Male/Female)	94/60	67/79	$\chi^2 = 6.9$	.009	25/13	45/45	$\chi^2 = 2.6$	.101	57/35	12/13	$\chi^2 = .81$	.368
Education, mean (SD)	13.9 (2.4)	15.1 (7.5)	$t = 1.85$	.064	13.9 (2.2)	15.8 (9.3)	$t = 1.22$	.224	14.0 (2.6)	13.5 (2.3)	$t = -.838$	.404
Educational years repeated, mean (SD)	1.8 (2.5)	2.3 (2.7)	$t = 1.47$	.141	.83 (.87)	1.1 (1.8)	$t = 1.00$	.319	3.1 (4.8)	5.0 (5.4)	$t = 1.70$	.091

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

| Having partnership most of the time in the year before study inclusion, No. (%) | 72 (24.7) | 85 (29.2) | $\chi^2 = 4.1$ | .036 | 17 (13.4) | 51 (40.2) | $\chi^2 = 1.6$ | .194 | 48 (41.4) | 18 (15.5) | $\chi^2 = 2.9$ | .085 |
| Population density in living area, mean (SD), habitants/km2 | 3717.8 (2532.3) | 3529.8 (2377.1) | $t = -.544$ | .587 | 4498.3 (2724.6) | 3022.1 (2274.3) | $t = -2.640$ | .010 | 3447.9 (2338.5) | 5152.1 (2361.8) | $t = 2.547$ | .013 |

17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 2. Classification Performance of the Clinical/Neurocognitive, GMV, and Combined Models and Validation Performance.

	True Positive, No.	True Negative, No.	False Positive, No.	False Negative, No.	Balanced Accuracy, %	AUC	Model P Value
Clinical/Neurocognitive Pure Model ROP-ROD	29	74	16	9	79.3	0.86	<0.001
Applied Clinical/Neurocognitive Model ROP+D-ROD+P	36	19	6	56	57.6	0.71	NA
Applied Clinical and Neurocognitive Model Validation Pure ROP-ROD	34	39	14	16	70.8	0.78	NA
Applied Clinical and Neurocognitive Mode Validation ROP+D-ROD+P	27	20	5	63	55	0.56	NA
GMV Pure Model ROP-ROD	15	77	13	23	62.5	0.7	<0.001
Applied GMV Model ROP+D-ROD+P	19	20	5	73	50.3	0.47	NA
Applied GMV Model Validation Pure ROP-ROD	17	38	15	33	52.8	0.59	NA
Applied GMV Model Validation ROP+D-ROD+P	31	17	8	59	51.2	0.6	NA
Combined Model ROP-ROD	30	72	18	8	79.5	0.87	NA
Applied Combined Model ROP+D-ROD+P	49	19	6	43	64.6	0.71	NA
Applied Combined Model Validation Pure ROP-ROD	36	38	15	14	71.8	0.78	NA

Applied Combined Model Validation ROP+D-ROD+P	31	18	7	59	53.2	0.56	NA

Transdiagnostic Features, Comorbidity and Classification of Recent Onset Psychosis and Depression: a Multimodal Machine Learning Approach

Paris Alexandros Lalousis, MSc[1,2]; Stephen J. Wood, PhD[1,3,4]; Lianne Schmaal, PhD[3,4]; Katharine Chisholm, PhD[1,5]; Sian Lowri Griffiths, PhD[1,2]; Renate L.E.P Reniers, PhD[1,2,6]; Alessandro Bertolino, MD[7]; Stefan Borgwardt, MD[8]; Paolo Brambilla, MD[9,10]; Joseph Kambeitz, MD[11]; Rebekka Lencer, MD, PhD[12]; Christos Pantelis, MB BS, MD, MRCPsych, FRANZCP[13]; Stephan Ruhrmann, MD[14]; Raimo K. R. Salokangas, MD, PhD, MSc[15]; Frauke Schultze-Lutter, PhD[16]; Carolina Bonivento, PhD[17]; Dominic Dwyer, PhD[11]; Adele Ferro, PsyD, PhD[10]; Theresa Haidl, MD[14]; Marlene Rosen, MSc[14]; Andre Schmidt, PhD[8]; Eva Meisenzahl, MD[16]; Nikolaos Koutsouleris, MD[11]*; Rachel Upthegrove, MBBS FRCPsych, PhD[1,2,18]* and the PRONIA Consortium#

* Joint Senior Authors

#See PRONIA consortium author list

Author Affiliations:

1 -Institute for Mental Health, University of Birmingham, Birmingham, United Kingdom

2 -Centre for Human Brain Health, University of Birmingham, Birmingham, United Kingdom

3 -Orygen, the National Centre of Excellence in Youth Mental Health; Melbourne, Australia

4 -Centre for Youth Mental Health, The University of Melbourne, Parkville, Australia

5 -Department of Psychology, Aston University, United Kingdom

6 –Institute of Clinical Sciences, University of Birmingham, United Kingdom

7 -Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari Aldo Moro, Bari, Italy

8 -Department of Psychiatry, University of Basel, Basel, Switzerland

9 -Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy

10 - Department of Neurosciences and Mental Health, IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy.

11 -Department of Psychiatry and Psychotherapy, Ludwig Maxmilians University, Munich, Germany

12 -Department of Psychiatry, University of Münster, Münster, Germany

13 -Melbourne Neuropsychiatry Centre, University of Melbourne, Melbourne, Australia

14 -Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital, University of Cologne, Cologne, Germany

35   15 -Department of Psychiatry, University of Turku, Turku, Finland

36   16 -Department of Psychiatry and Psychotherapy, Heinrich-Heine University Düsseldorf,
37   Düsseldorf, Germany

38   17 -IRCCS "E. Medea" Scientific Institute, San Vito al Tagliamento (Pn), Italy

39   18 -Early Intervention Service, Birmingham Women's and Children's NHS trust, United
40   Kingdom

41

42

43

44

45

46

47

48

49

50

51

52

54

55   **Corresponding Author**:

56   Mr. Paris Alexandros Lalousis, BSc, MSc

57   School of Psychology

58   University of Birmingham, 52 Pritchatts Road

59   B15 2SA, Birmingham, United Kingdom

60   pal532@student.bham.ac.uk

61

62   *Supplementary Methods and Results*

Lalousis et al.   Transdiagnostic features, comorbidity and classification

63 **1. Supplementary Methods**

64 **1.1. PRONIA recruitment infrastructure**

65 The 300 study participants (154 individuals with ROP and 146 individuals with ROD) analyzed in

66 the present study were recruited following a standardized recruitment and ascertainment protocol

67 (see **Figure S1** and **Table S3**). The observational study protocol involved follow-up examinations

68 every three months after the index ascertainment and was implemented by the following 7 PRONIA

69 sites:

70 **Table S1: Characteristics of the recruiting institutions in the PRONIA consortium.**

PRONIA Site	Institution Name	Country	Type of Service	Catchment Population	Screening population / year
Munich	Department of Psychiatry and Psychotherapy, Ludwig-Maximilian-University Munich	DE	Academic outpatient services including specialized service for early recognition of psychosis; tertiary care academic hospital	1,200,000	700
Basel	Department of Psychiatry and Psychotherapy, University of Basel	CH	Academic inpatient and outpatient services including specialized service for early recognition and intervention of psychosis; tertiary care academic hospital	500,000	200
Milan Niguarda	Department of Pathophysiology and Transplantation, University of Milan. Four recruitment hospitals: Niguarda, Policlinico, San Paolo, Villa San Benedetto Menni in Albese con Cassano	IT	Psychiatric outpatient services including specialized services for early recognition of psychosis and persons at high risk; Academic hospital, providing psychiatric inpatient services, psychiatric outpatient services and local services;	600,000	1,000
Cologne	Department of Psychiatry and Psychotherapy, University of Cologne	DE	Academic outpatient services including specialized service for early recognition of psychosis; tertiary care academic hospital	1,000,000	600
Birmingham	The University of Birmingham	UK	Academic specialised Early Intervention Service for Psychosis covering Birmingham and Solihull. Community and Inpatient	1,200,000	800
Turku	Department of Psychiatry, University of Turku	FI	Psychiatric outpatient and hospital services responsible for treatment of psychiatric patients in their catchment areas in the South-Western Finland.	284,000	2,300

Lalousis et al.

Transdiagnostic features, comorbidity and classification

Udine	Department of Psychiatry, University of Udine	IT	Psychiatic outpatient services, academic hospital and local services. Tertiary care neuropsychiatric service	600,000	500

71 Upon study enrolment, the participants were pseudonymized twice, locally at each site and

72 centrally at the level of the PRONIA portal. The PRONIA portal consists of a multi-user database

73 hosting the clinical and neurocognitive information, and defaced MR images obtained from the

74 study participants. The data are organized into digital questionnaires, visits, and cases. The portal

75 provides the case managers with a controlled web-based interface to enter and upload the different

76 data into the respective questionnaires. Furthermore, the PRONIA consortium has implemented a

77 PRONIA@home mobile device interface that allows the study participants to securely log into the

78 portal and fill out the self-rating questionnaires of given visit. Upon completion of the data entry

79 across all questionnaires of given visit, the data are checked by an automatic quality control

80 procedure which executes approximately 1600 data integrity and dependency rules. These rules

81 include (1) basic checking of missing data and data ranges, (2) checking of dependency within one

82 questionnaire, (3) dependencies between two questionnaires within one visit, and (4) dependencies

83 between two consecutive visits (such as consistency of dates). Detected errors are fed back to the

84 respective case managers allowing for a manual correction of the respective issues. This process is

85 re-iterated until the quality of the clinical questionnaires in the given visit is sufficient for the entire

86 visit to be locked.

Lalousis et al.   Transdiagnostic features, comorbidity and classification

1.2. PRONIA study design and examination instruments

A comprehensive battery of ascertainment tools was used within a longitudinal observational study design to generate a multi-modal phenotypic profile of each study participant (see **Figure S1** and **Table S2**). The clinical part of the battery compiled questionnaires that capture sociodemographic, somatic, environmental, diagnostic, psychopathological, functional and quality-of-life related variables in the PRONIA study population. This battery was complemented by multi-domain neurocognitive and neuroimaging examinations as well as blood sampling for later genetic characterization, which were carried out at the baseline and 9-month follow-up timepoints.

Table S2: Clinical and neurocognitive examinations performed in the CHR, ROD, ROP, and HC groups during the 18-month follow-up period of the study. **Clinical assessment types:** *OR* Observer-based rating instrument, *SR* Self-rating-based instrument. **Examination timepoints:** *T0* Baseline examination, *IV3/IV6/IV12/IV15* 3, 6, 12, 15-month examinations conducted only in the clinical study participants, *T1* 9-month examination, *T2* 18-month follow-up examination. **Observer-based instruments:** CAARMS Comprehensive Assessment of the At-Risk Mental States[1], *CHR Criteria* Clinical High-Risk criteria summary questionnaire, *FROGS* Functional Remission in General
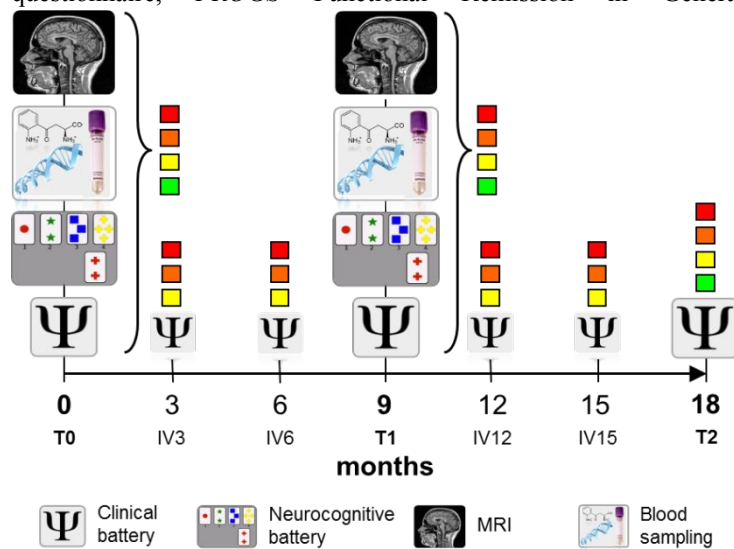
Figure S1: Observational study design of PRONIA. Colored boxed indicate type of assessment / visits conducted in each of the study groups: Healthy controls (green), patients with recent-onset depression (yellow), persons with a clinical high-risk for psychosis (orange), patients with recent-onset psychosis (red).

Schizophrenia[2], *GAF* Global Assessment of Functioning, *GF:S/R* Global Functioning: Social / Role[3], *PANSS* Positive and Negative Symptom Scale[4], *PAS* Premorbid Adjustment Scale[5], *SANS* Scale for the Assessment of Negative Symptoms[6], *SCID-IV Screening/Summary* Structured Clinical Interview for DSM-IV[7], *SIPS* Standardized Interview for the Assessment of Prodromal Symptoms (modified version 5.0)[8], *SPI-A [COGDIS/COPER]* Schizophrenia Proneness Instrument [Cognitive disturbances (COGDIS) / Cognitive-Perceptual (COPER) disturbances][9], Transition Criteria Interval questionnaire for the assessment of transition criteria, *UHR - Schizotypy, Genetic Risk* Interview for the Assessment of Schizotypal personality traits, and familial risk for psychosis. **Self-rating instruments:** *BDI-II* Beck Depression Inventory II[10], *CISS-24* Coping Inventory for Stressful Situations – 24 items[11], *CTQ* Childhood Trauma Questionnaire[12], *EHI-SR* Edinburgh Handedness Inventory – Short Version[13], *EDS* Everyday Discrimination Scale – Modified Version[14], *LEE* Level of Expressed Emotions[15], *MSPSS* the Multidimensional Scale for Perceived Social Support[16], *NEO-FFI* NEO Five Factor Inventory of Personality Traits[17], *RSA* Resilience Scale for Adults[18], *SPIN* Social Phobia Inventory[19], *WHO-QOL-BREF* WHO Quality of Life Questionnaire-Brief Version[20]. Neurocognitive tests: *CPT-IP* Continuous-Performance Test-Identical Pairs (adapted tablet version)[21], *DANVA* Diagnostic Analysis of Non-Verbal Accuracy 2 (adapted tablet version)[22], *DS* Auditory Digit Span (Forward/Backward) adapted from the PEBL battery, *DSST* Digit-Symbol-Substitution Test from the BACS battery, *ROCF* Rey-Osterrieth complex figure[23], *SAT*

Lalousis et al.　　　　　Transdiagnostic features, comorbidity and classification

124 Salience Attribution Task (adapted version)[24], *SOPT* self-ordered pointing task (adapted version)[25], *TMT-A/-B* Trail-
125 Making Test A and B[26], *VF* phonemic/semantic verbal fluency test.

Instrument	Form	Screening		T0		IV3	IV6	T1		IV12	IV15	T2	
		PAT	HC	PAT	HC	PAT	PAT	PAT	HC	PAT	PAT	PAT	HC
General Data	OR	X	X					X	X			X	X
Reasons for Referral	OR	X											
Treatment Documentation	OR	X	X			X	X	X	X	X	X	X	X
Somatic state and Health History	OR	X	X					X	X			X	X
SPI-A COGDIS/COPER	OR	X	X			X	X	X	X	X	X	X	X
SIPS positive symptoms	OR	X	X			X		X	X			X	X
CAARMS	OR	X	X			X		X	X			X	X
GAF	OR	X	X			X		X	X			X	X
UHR – Schizotypy, Genetic Risk	OR	X	X			X		X	X			X	X
CHR Criteria	OR	X	X					X	X			X	X
Transition Criteria	OR					X	X			X	X		
SCID-IV Screening	OR	X	X					X	X			X	X
SCID-IV Summary	OR	X	X					X	X			X	X
Demographic and Biographic Data	OR			X	X			X	X			X	X
PAS	OR			X	X			X				X	
SPI-A	OR			X	X			X				X	
SIPS negative, disorganized and general symptoms	OR			X	X			X				X	
PANSS	OR			X		X	X	X		X	X	X	
SANS	OR			X				X				X	
Chart of Life Events	OR			X	X	X	X	X	X	X	X	X	X
FROGS	OR			X				X				X	
GF: Social & Role	OR			X	X	X	X	X	X	X	X	X	X
Prognostic evaluation	OR			X				X				X	
MSPSS	SR			X	X			X	X			X	X
RSA	SR			X	X			X	X			X	X
CISS 24	SR			X	X			X	X			X	X
SPIN	SR			X	X			X	X			X	X
BDI-II	SR			X	X	X	X	X	X	X	X	X	X
WHO-QOL-BREF	SR			X	X			X	X			X	X
EHI-SR	SR			X	X								
LEE	SR			X	X			X	X			X	X
Wisconsin Scales	SR			X	X								
EDS	SR			X	X								
Bullying Scale T0	SR			X	X								
CTQ	SR			X	X								
NEO-FFI	SR			X	X								
DS backward (BACS)	NPT			X	X			X	X				
DS forward (BACS)	NPT			X	X			X	X				
CPT-IP (BACS)	NPT			X	X			X	X				
DANVA	NPT			X	X			X	X				
DSST	NPT			X	X			X	X				
RAVLT	NPT			X	X			X	X				
ROCF	NPT			X	X			X	X				
SAT	NPT			X	X			X	X				
SOPT	NPT			X	X			X	X				
TMT-A	NPT			X	X			X	X				
TMT-B	NPT			X	X			X	X				
VF phonetic	NPT			X	X			X	X				
VF semantic	NPT			X	X			X	X				
WAIS-III	NPT			X	X			X	X				

Lalousis et al.   Transdiagnostic features, comorbidity and classification

126   **1.3.   MRI harmonization and data acquisition**

127   When setting up the PRONIA study, we decided to generate a MRI database that would represent

128   the MR scanner sequence heterogeneity encountered in clinical real-world. The aim of this strategy

129   was to strengthen the generalizability and clinical applicability of the predictive models developed

130   by our machine learning analyses. Thus, we agreed on a minimal harmonization protocol that

131   required the PRONIA sites to only (1) acquire isotropic or nearly isotropic voxel sizes of preferably

132   1 mm resolution, (2) set the Field Of View (FOV) parameters accordingly to guarantee the full 3D

133   coverage of the brain including all parts of the cerebellum, and (3) define the relaxation time (TR)

134   and echo time (TE) as well as other imaging parameters in a way that would maximize the contrast

135   between cortical ribbon and the white matter and enhance the signal-to-noise ratio in the images.

136   At every site all the images were visually inspected, automatically defaced and anonymized using

137   an in-house Freesurfer-based script before the data was centralized. **Table S3** lists the parameters

138   defining the structural MR sequences used to examine in the PRONIA discovery sample

139   participants.

140   **Table S3: MR scanner systems and structural MRI sequence parameters** used at the
141   respective PRONIA sites.

PRONIA Site	Model	Field Strength	Coil Channels	Flip Angle	TR [ms]	TE [ms]	Voxel Size [mm]	FOV	Slice Number
Munich	Philips Ingenia	3T	32	8	9.5	5.5	0.97 x 0.97 x 1.0	250 x 250	190
Milan Niguarda	Philips Achieva Intera	1.5T	8	12	Shortest (8.1)	Shortest (3.7)	0.93 x 0.93 x 1.0	240 x 240	170
Basel	SIEMENS Verio	3T	12	8	2000	3.4	1.0 x 1.0 x 1.0	256 x 256	176
Cologne	Philips Achieva	3T	8	8	9.5	5.5	0.97 x 0.97 x 1.0	250 x 250	190
Birmingham	Philips Achieva	3T	32	8	8.4	3.8	1.0 x 1.0 x 1.0	288 x 288	175
Turku	Philips Ingenuity	3T	32	7	8.1	3.7	1.0 x 1.0 x 1.0	256 x 256	176
Udine	Philips Achieva	3T	8	12	Shortest (8.1)	Shortest (3.7)	0.93 x 0.93 x 1.0	240 x 240	170

142 **1.4. MRI processing pipeline**

143 The manual of the CAT12 toolbox (http://www.neuro.uni-jena.de/cat12/CAT12-Manual.pdf)

144 details the processing steps applied to the structural images. These steps consist of:

145 (1) A 1st denoising step based on Spatially Adaptive Non-Local Means (SANLM) filtering[27].

146 (2) An Adaptive Maximum A Posteriori (AMAP) segmentation technique, which models local

147 variations of intensity distributions as slowly varying spatial functions and thus achieves a

148 homogeneous segmentation across cortical and subcortical structures[28].

149 (3) A 2nd denoising step using Markov Random Field approach which incorporates spatial prior

150 information of adjacent voxels into the segmentation estimation generated by AMAP[28].

151 (4) A Local Adaptive Segmentation (LAS) step, which adjusts the images for white matter (WM)

152 inhomogeneities and varying gray matter (GM) intensities caused by differing iron content in

153 e.g. cortical and subcortical structures. The LAS step is carried out before the final AMAP

154 segmentation.

155 (5) A Partial Volume Segmentation algorithm that is capable of modeling tissues with intensities

156 between GM and WM, as well as GM and cerebrospinal fluid (CSF) and is applied to the

157 AMAP-generated tissue segments.

158 (6) A high-dimensional DARTEL registration of the image to a MNI-template generated from the

159 MRI data of 555 healthy controls in the IXI database (http://www.braindevelopment.org).

160 (7) The GM maps were then multiplied with the Jacobian determinants that were obtained during

161 registration in order to produce GM volume maps.

162 (8) The Quality Assurance framework of CAT12 was used to check the quality of the GMV maps.

163 **1.5. Support Vector Machine Learning Preprocessing**

Lalousis et al.   Transdiagnostic features, comorbidity and classification

164   Prior to the machine learning analysis the following preprocessing steps were completed:

165   a) every feature was scaled from 0 to 1 and completely non-finite features were zeroed-out; b)

166   any feature that had infinite values was pruned; c) missing values were imputed using KNN

167   Euclidean distance median replacement using 7 nearest neighbours. For each missing value of a

168   given CV1 or CV2 subject, a subset of cases that had values for the given variable and had values

169   in all other variables which were non-empty were identified. Subjects in the source subset were

170   sorted according to their similarity with the target subject using the Euclidean distance. Then, the

171   median of the given variable was computed using the 7 nearest neighbours. The original, non-

172   imputed training matrix was used at all times d) site effects were corrected for. This step

173   consisted of i) performing a principal components analysis, ii) identifying the components that

174   are most correlated with site variance, and iii) reconstructing the data without these components

175   in order to effectively remove such effects from the data. Any component with a Spearman

176   correlation coefficient greater than 0.5 was excluded from the analysis and 90% of the variance

177   was retained. e) the data was scaled again. The model performance criterion that was used was

178   balanced accuracy and the learning algorithm that was chosen was LIBSVM. Imbalanced

179   learning was corrected for by increasing the C value in the minority class by multiplying it by the

180   inverse ratio of the training class sizes (weighting the hyperplane). The kernel type that was

181   linear with eleven learning parameters in order to optimize the choice of C value. Wrapper

182   methods were activated at all parameter combinations with greedy sequential backward feature

183   selection (Stop at k=90% of features; Feature stepping at 10% of worst performing features at

184   each cycle). Starting with the full feature set the SVM was ran iteratively with the 10% worst

185   performing features being eliminated at each cycle until 10% of the feature pool was left. In order

186   to assess the models' statistical significance we performed permutation analysis to create an

Lalousis et al.   Transdiagnostic features, comorbidity and classification

187   empirical null distribution of weights for each feature and then compare the observed weight to

188   this distribution. The models were retrained in the cross-validation framework using the

189   respective feature and label subsets obtained from the observed-label analyses 1000 times. For

190   each permutation the predictions were accumulated into a permuted ensemble prediction for each

191   CV2 subject. In that way a null distribution of out-of-training classification performance for the

192   prediction models was produced. The significance of the observed out-of-training classification

193   performance was calculated as the number of events where the permuted out-of-training

194   classification performance was higher or equal to the observed classification performance divided

195   by the number of permutations performed. Then the significance of the model was determined

196   according to a p threshold of p<0.05. Furthermore we applied a sign-based consistency algorithm

197   to calculate the number of times that the sign of each feature (positive or negative) was consistent

198   within an ensemble multiplied by the number of times that the feature was non-zero. The measure

199   is between 0 to 1, with 1 representing perfect consistency within the ensemble and 0 if the

200   weights are equally positive and negative or when the feature is omitted with a zero weight. A p-

201   value was then calculated by defining a hypothesis test for the importance score with a null

202   hypothesis of 0. A z-score was calculated as the importance divided by the square root of the

203   variance of the importance scores. A standard p-value was then calculated using a normal

204   cumulative distribution function to choose the right-tailed significance. P-values were then

205   corrected using the false-discovery rate.

206   For the neuroimaging model the same steps used apart from: 1) when inputting the grey matter

207   volume (GMV) images into the modality a G-theory mask[29] was used for quantifying the degree

208   of reliability for the imaging modalities based on a travelling participants study and 2) during

209   preprocessing of the neuroimaging model the following additional steps were completed a)any

Lalousis et al.   Transdiagnostic features, comorbidity and classification

210   feature that had infinite values was pruned; b) features were ranked up using external ranking by

211   means of hard selection thresholds (15%, 25%, and 50%). Only the top 15%, 25%, and 50%

212   features as defined by the G-theory mask were kept; c) Principal Component Analysis (PCA)

213   Dimensionality Reduction (dimensions: 0.8); d) every feature was scaled from 0 to 1 and

214   completely non-finite features were zeroed out; 3) when visualizing the classification a 8 FWHM

215   kernel was used to smooth the images.

216   Finally, a stacking-based data fusion framework[30,31] was used to examine whether the

217   combination of the clinical-based and the neuroimaging-based models would provide a superior

218   classification accuracy. To achieve this, the decision scores of the neuroimaging and the clinical

219   models were combined, standardized, and forwarded to a greedy sequential forward search

220   algorithm[32] which found a parsimonious combination of classifiers maximizing $\overline{PSI_{reg}}$ across the

221   C parameter range by employing L2-regularized logistic regression (L2LR)[33]. Each L2LR

222   ensemble was then applied to the standardized neuroimaging and clinical based decision scores

223   available for the CV_2 validation data. Class prediction was achieved by using majority voting on

224   Majority voting on P_{ens}.

1.6. Hyperparameter Optimization

226   For hyperparameter optimization, we computed

227    $\overline{BAC_{reg}} = \Sigma_{i=1} \left(\frac{n_{TP_i}}{n_{TP_i} + n_{FN_i}} + \frac{n_{TN_i}}{n_{TN_i} + n_{FP_i}} \right) \Big/ 2$ at given parameter combination across all k CV_1

228   partitions with $n_{TP_i}/n_{TP_i} + n_{FN_i}$ being Sensitivity and $n_{TN_i}/n_{TN_i} + n_{FP_i}$ being Specificity, and the

229   fraction of the training population serving as support vectors in the ith CV_1 partition. Our

230   optimization technique's aim was to find a combination of T_G, PCs and the SVM's regularization

Lalousis et al.   Transdiagnostic features, comorbidity and classification

231   parameter C [range: $2^{[-3 \underset{\in \mathbb{Z}}{\rightarrow} +4]}$ that maximized $\overline{\text{BAC}_{\text{reg}}}$ within a 3 (T_G) × 5 (PC) × 8 (C)

232   hyperparameter cube. The optimized ensemble was then applied to the CV_2 validation data to

233   produce a mean decision score ($\overline{D_{ens}}$) and majority voting-based class membership probabilities (

234   P_{ens}) for each CV_2 validation subject. This produced a mean decision score ($\overline{(D\_ens)}$) and

235   majority voting-based class membership probabilities (P_ens) for each CV2 validation subject.

1.7. Stacking Fusion Framework

237   The decision scores of the neuroimaging and the clinical models were combined,

238   standardized, and forwarded to a greedy sequential forward search algorithm[32] which found a

239   parsimonious combination of classifiers maximizing $\overline{\text{PSI}_{\text{reg}}}$ across the C parameter range by

240   employing L2-regularized logistic regression (L2LR)[33]. Each L2LR ensemble was then applied to

241   the standardized neuroimaging and clinical based decision scores available for the CV_2 validation

242   data. Class prediction was achieved by using majority voting on Majority voting on P_{ens}.

1.8. Clinical/Neurocognitive Model List of Tests

244   The clinical and neurocognitive tests used to train the clinical support vector machine

245   learning model were the following: (1) Wisconsin Schizotypy Scale (WSS) physical and social

246   anhedonia subscales[34–36]; (2) Premorbid Adjustment Scale (PAS)[37,38], (3) Functional Recovery

247   Scale in Schizophrenia (FROGS)[2]; (4) SIPS-negative[39]; (5) Schizophrenia Proneness Instrument,

248   Adult version (SPI-A)[40]; (6) Resilience Scale for Adults (RSA)[18,41,42]; and a range of

249   neurocognitive tests: (1) Digit Span Test (DST)[43,44]; (2) Phonemic Verbal Fluency (PVF)[43,45]; (3)

250   Semantic Verbal Fluency (SVF)[43,45]; (4) Auditory-Verbal Learning Test (AVLT)[43], and (5) Digit

251   Symbol Substitution Test (DSST)[43,44].

Lalousis et al.   Diagnosis and Classification in Recent Onset Psychosis and Depression

Table S4: DGPPN S3 Guidelines for the treatment of first-episode psychosis and schizophrenia *(translated English version of Table 4.1 stated in the short version of the guideline manual available in* [*https://www.dgppn.de/_Resources/Persistent/a6e04aa47e146de9e159fd2ca1e6987853a055d7/S3_Schizo_Kurzversion.pdf*](https://www.dgppn.de/_Resources/Persistent/a6e04aa47e146de9e159fd2ca1e6987853a055d7/S3_Schizo_Kurzversion.pdf)). *Candidate CHR and ROD patients were excluded if they had received antipsychotic medication (1) for more than 30 cumulative days at or above the minimum target dosage threshold for the treatment of first-episode psychosis, or (2) within the past 3 months before psychopathological baseline assessments at or above the minimum target dosage threshold for the treatment of first-episode psychosis. Abbreviations:* DI *dosage interval, [2]maximum recommended dosage according to prescribing information.*

Substance	Recommended starting dosage (mg/d)	DI'	Target dosage first-episode psychosis (mg/d)	Target dosage relapsing schizophrenia (mg/d)	Maximum dosage recommended (mg/d)[2]
Atypical Antipsychotics					
Amisulpride	200	(1)-2	100-300	400-800	1200
Aripiprazole	(10)-15	1	15-(30)	15-30	30
Olanzapine	5-10	1	5-15	5-20	20
Quetiapine	50	2	300-600	400-750	750
Risperidone	2	1-2	1-4	3-6-(10)	16
Ziprasidone	40	2	40-80	80-160	160
Typical Antipsychotics					
Fluphenazine	0.4-10	2-3	2.4-10	10-20	20-(40)
Flupentixole	2-10	1-3	2-10	10-60	60
Haloperidole	1-10	(1)-2	1-4	3-15	100
Perazine	50-150	1-2	100-300	200-600	1000
Perphenazine	4-24	1-3	6-36	12-42	56
Pimozide	1-4	2	1-4	2-12	16
Zotepine	25-50	2-(4)	50-150	75-150	450
Zuclopenthixole	2-50	1-3	2-10	25-50	75

Lalousis et al.

Transdiagnostic features, comorbidity and classification

264

Table S3: SCID DSM-IV diagnoses detailed breakdown in ROP groups (Discovery Sample)

	ROP total N=154
DSM Diagnosis	N (%)
Schizophrenia	72 (46.7)
Schizophreniform	13 (8.0)
Psychosis NOS	18 (11.7)
Delusional Disorder	11 (7.1)
Brief psychotic disorder	6 (3.8)
Substance Induced	3 (1.9)
Schizoaffective Disorder	10 (6.5)
Bipolar Disorder	12 (7.8)
Severe depression with psychosis	9 (5.7)

266

267

Lalousis et al.   Transdiagnostic features, comorbidity and classification

268   **Table S4: Demographics of Replication Sample**

	ROP Group	ROD Group	t/χ2	P Value	Pure ROP Group	Pure ROD Group	t/χ2	P Value	ROP+D Group	ROD+P Group	t/χ2	P Value
Sample Sizes, No.	161	131			50	53			90	25		
Age, Mean (SD)	25.6 (6.1)	25.0 (5.9)	t =.787	.432	25.8 (6.3)	25.7 (6.1)	$t =$ .039	.969	25.3 (6.1)	25.5 (6.5)	$t = -$ .156	.876
Sex (Male/Female)	89/71	68/62	χ^2 =.339	.844	30/20	24/29	χ^2 =2.234	.135	48/42	15/10	χ^2 =.351	.554

269

270

271

272

273

274

275

276

15

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

2. Supplementary Results

A number of exploratory supplementary analyses are presented in the supplementary results. These include a) correlational analyses between the clinical and imaging decision scores of the four different groups (Pure ROP/Pure ROD/ROP+D/ROD+P), in order to explore whether there was an association between clinical, and neuroanatomical features driving misclassification; b) association and comparison analyses between patients who were correctly classified and patients who were misclassified, to explore any potential 'sub-group' identification of misclassified participants; c) whole brain GMV comparison between the whole ROP and the whole ROD group, between subjects who were classified as ROP and subjects who were classified as ROD; and d) group level comparisons of the decisions scores of the three machine learning models.

2.1. Correlation of Clinical and Imaging Decision scores.

We wanted to examine the relationship between the clinical and imaging decision scores of our different groups (Pure ROP/Pure ROD/ROP+D/ROD+P) employing a Pearson correlation test. In figure S2 the non-significant positive correlation between the clinical and imaging decision scores in the Pure ROP group can be seen ($r = 0.005$, $p<.05$) In figure S3 we can see the positive correlation between the clinical and imaging decision scores in the Pure ROD group ($r = .112$, $p=.073$). Figure S4 shows the positive correlation between the clinical and imaging decision scores in the ROP+D group ($r = .062$, $p <.05$). Finally, figure S5 shows the positive correlation between the clinical and imaging scores in the ROD+P group ($r = .387$, $p =0.14$).

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression
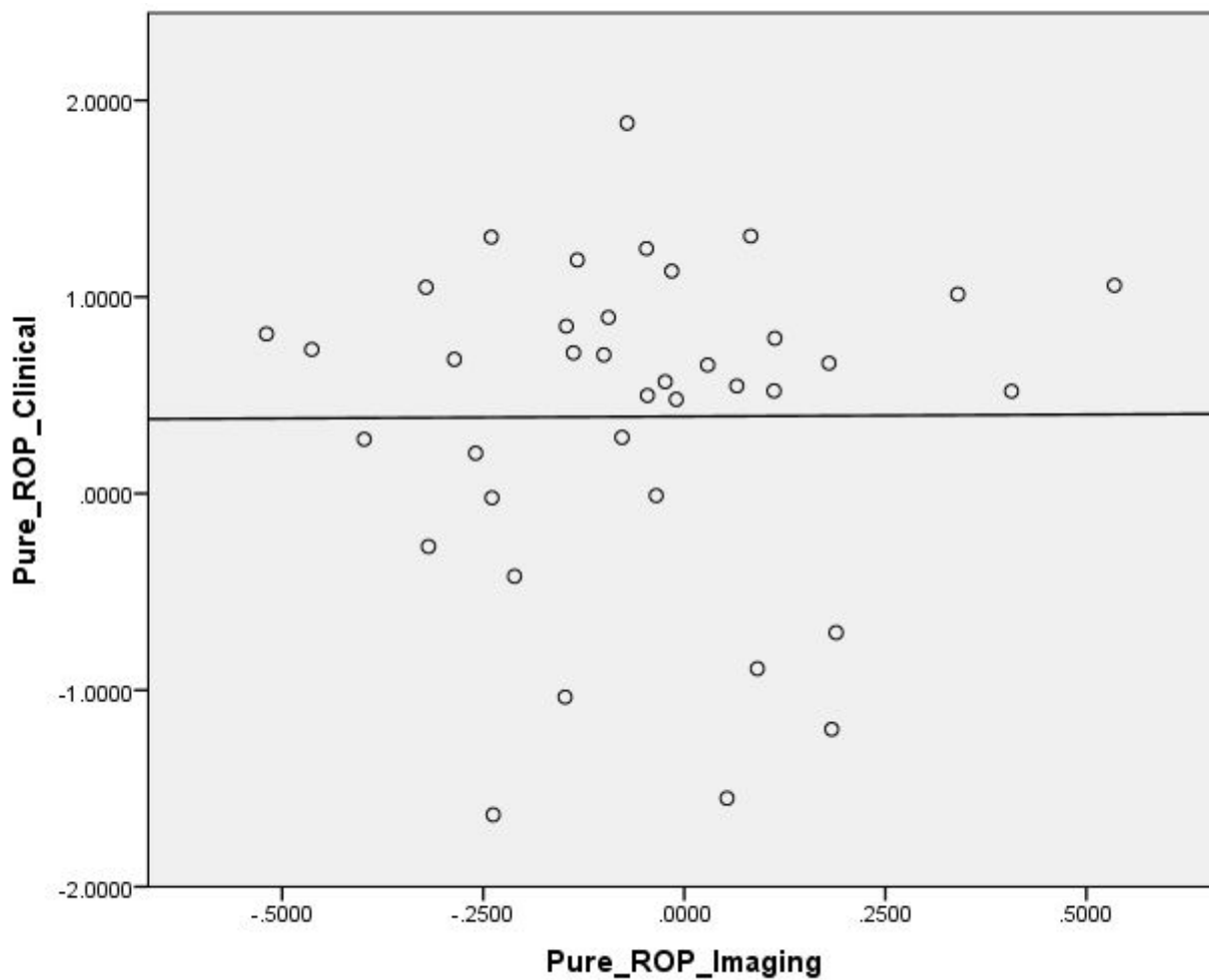

Figure S2. Pure ROP Clinical and Imaging Decision Scores Correlation

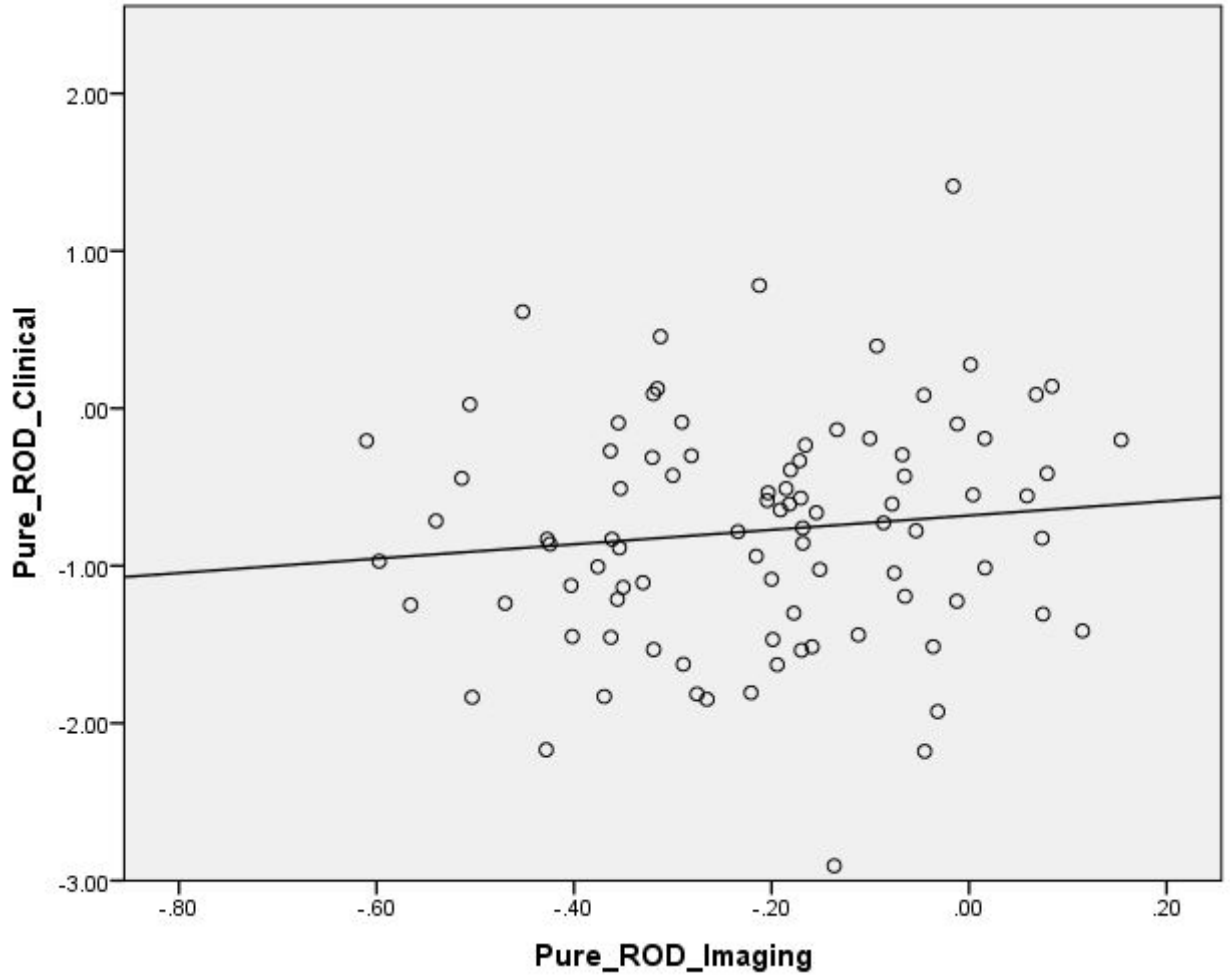Lalousis et al. Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

Figure S3. Pure ROD Clinical and Imaging Decision Scores Correlation

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression
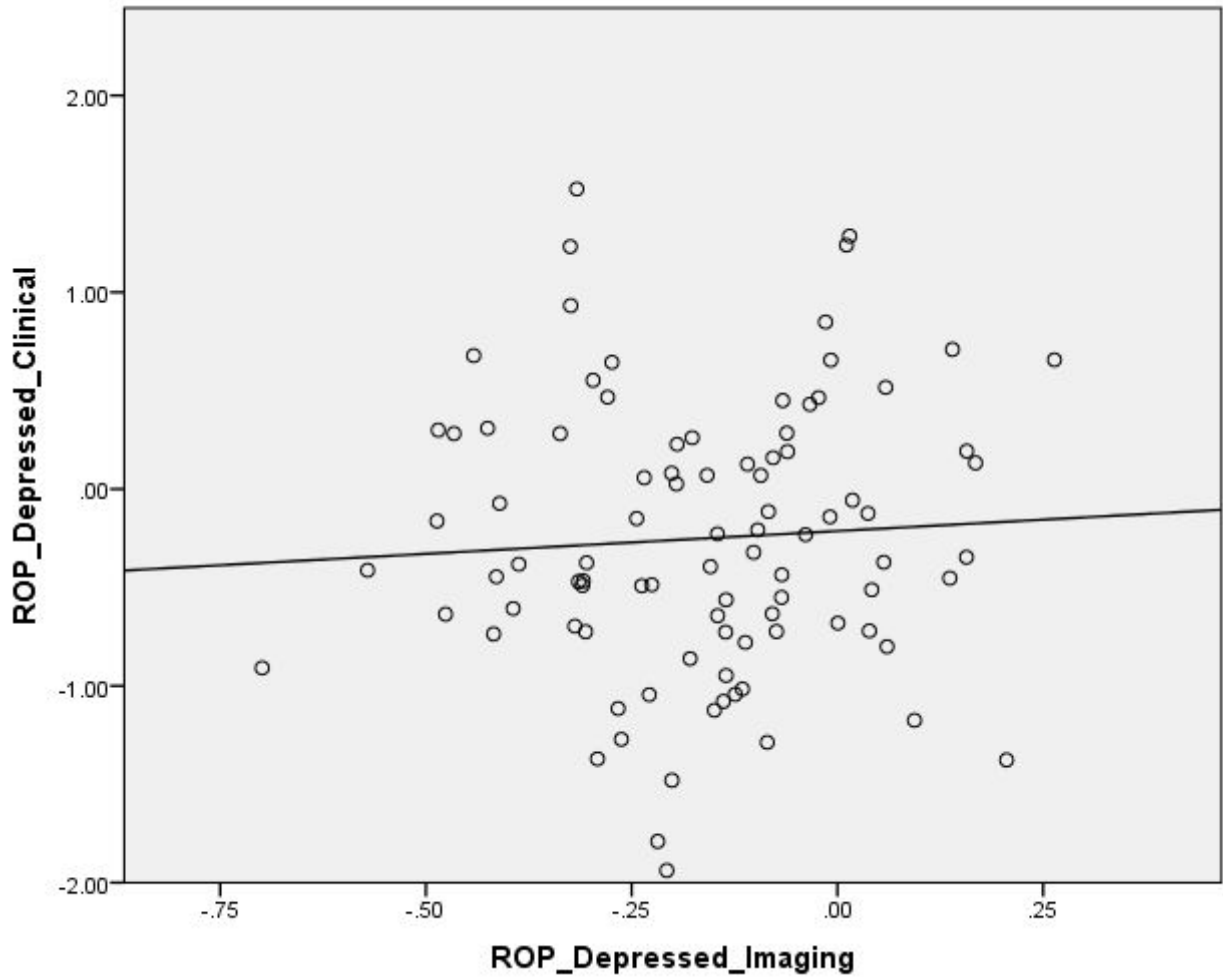

Figure S4. ROP Depressed Clinical and Imaging Decision Scores Correlation

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression
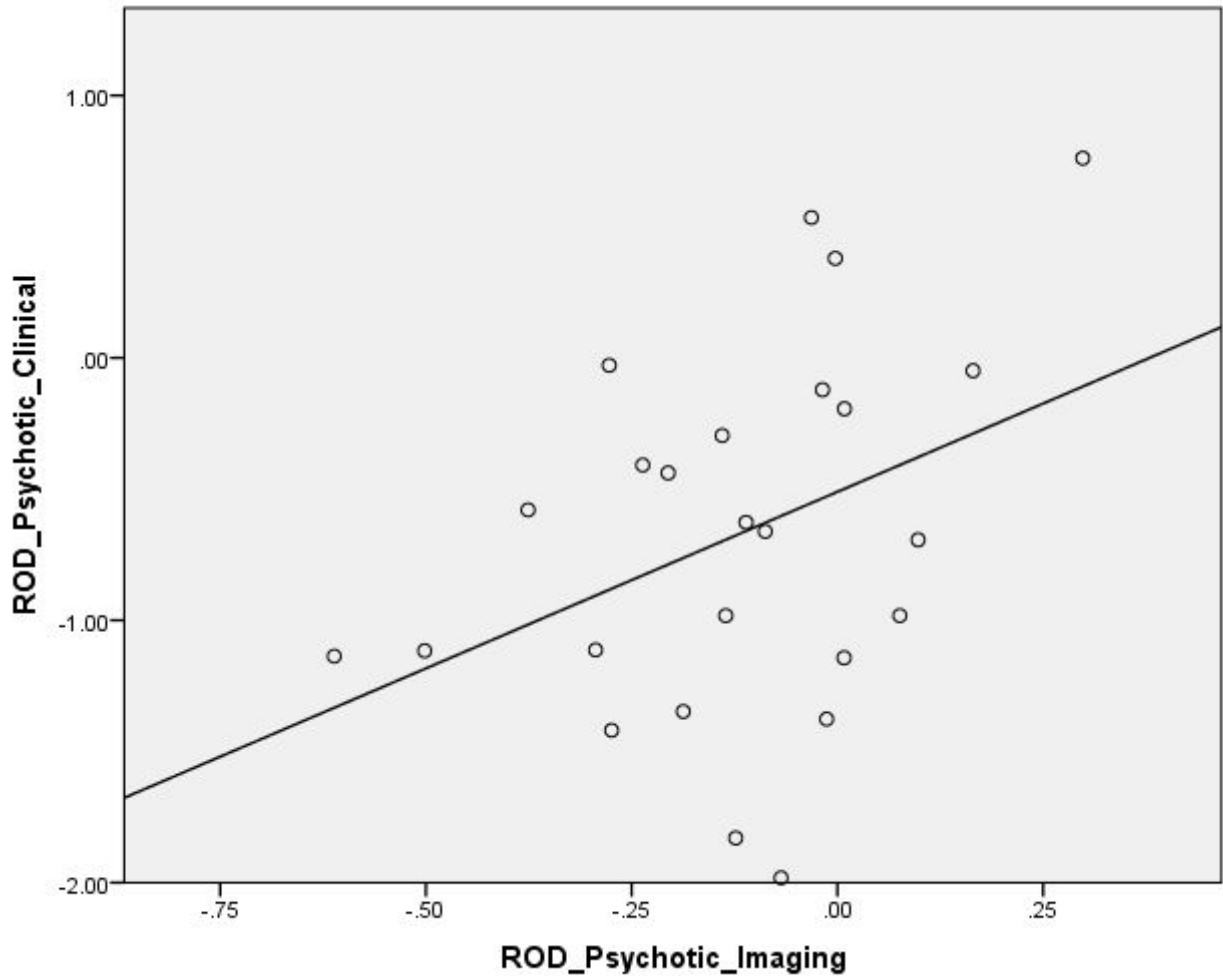

Figure S5. ROP Depressed Clinical and Imaging Decision Scores Correlation

2.2. Association and comparison analysis between patients who were correctly classified

and patients who were misclassified

To understand the reasons why some of the patients in our sample were misclassified we performed a series of association and comparison analysis between those patients who were correctly classified and those who were misclassified using clinically relevant measures. In table S5 we can see the association between Pure ROP patients who were correctly classified as ROP and Pure ROP patients who were misclassified as ROD both in the clinical and imaging models. In Table S6 we have compared Pure ROP patients who have been correctly classified and Pure ROP patients who have been misclassified in seven clinically relevant measures (PANSS

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

317 Positive, WSS Perceptual Aberration, CAARMS Unusual Thought Content, CAARMS Non-

318 Bizarre Ideas, and CAARMS Disorganized Speech), both in the clinical and the imaging models.

319 Furthermore we performed an association analysis between ROP+D patients who had been

320 correctly classified and ROD+D patients who had been misc**lassified (Table S7).** Correctly

321 classified ROP+D patients were compared to misclassified ROP+D patients in the same seven

322 measures that were mentioned previously (Table S8).

323

324

325

326

327
328 **Table S5. Association Analyses between Pure ROP Classification and Schizotypy Criteria**

		SPD Criteria Not Met Count	SPD Criteria Not Met Expected Count	SPD Criteria Met Count	SPD Criteria Met Expected Count	$\chi 2$	P Value
Clinical Classification	Correct Classification	23	23.8	3	2.2		
	Misclassification	10	9.2	0	0.8	1.25	.131
Imaging Classification	Correct Classification	12	12.8	2	1.2		
	Misclassification	21	20.2	1	1.8	1.06	.151

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

329

Table S6. Comparison Analyses between Pure ROP Correctly Classified and Pure ROP Misclassified (P-Values with an asterisk survive FDR (Benjamini-Hochberg) correction threshold)

		Correctly Classified Mean and Standard Deviation	Misclassified Mean and Standard Deviation	t	P Value
Clinical Classification	PANSS Positive Total	16 (5.3)	13.7 (8.0)	1.013	.159
	WSS Magical Ideation Total	2.0 (2.0)	3.5 (2.7)	-1.780	.041*
	WSS Perceptual Aberration Total	.5 (.9)	1.1 (1.8)	-1.208	.117
	CAARMS Unusual Thinking Content	4.8 (2.1)	5.2 (1.9)	-.485	.315
	CAARMS Non Bizarre Ideas	4.6 (2.1)	2.7 (2.9)	2.203	.017*
	CAARMS Perceptual Abnormalities	3.6 (2.7)	5.2 (1.4)	-1.740	.045*
	CAARMS Disorganized Speech	2.4 (2.0)	2.0 (2.4)	.544	.295
Imaging Classification	PANSS Positive Total	13.1 (4.5)	16.8 (6.6)	-1.888	.034
	WSS Magical Ideation Total	2.8 (2.4)	2.1 (2.1)	.965	.170
	WSS Perceptual Aberration Total	.4 (1.3)	.8 (1.2)	-.846	.201
	CAARMS Unusual Thinking Content	4.6 (2.1)	5.0 (2.0)	-.599	.276

330
331
332

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

CAARMS Non Bizarre Ideas	4.4 (2.3)	3.8 (2.5)	.724	.237
CAARMS Perceptual Abnormalities	4.6 (2.2)	3.6 (2.6)	1.123	.134
CAARMS Disorganized Speech	2.2 (1.9)	2.3 (2.2)	-.114	.455

333

334

335

336   **Table S7. Association Analyses between ROP Depressed Classification and Schizotypy**
337   **Criteria**

		SPD Criteria Not Met Count	SPD Criteria Not Met Expected Count	SPD Criteria Met Count	SPD Criteria Met Expected Count	χ2	P Value
Clinical Classification	Correct Classification	28	30.3	6	3.7		
	Misclassification	54	51.7	4	6.3	2.55	.055
Imaging Classification	Correct Classification	16	16	2	2		
	Misclassification	66	66	8	9	.001	.485

338   **Table S8. Comparison Analyses between ROP Depressed Correctly Classified and ROP**
339   **Depressed Misclassified (P-Values with an asterisk survive FDR (Benjamini-Hochberg)**
340   **correction threshold)**

	Correctly Classified Mean and Standard Deviation	Misclassified Mean and Standard Deviation

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

				t	P Value
Clinical Classification					
	PANSS Positive Total	20.0 (7.2)	17.9 (5.3)	1.609	.055*
	WSS Magical Ideation Total	5.1 (3.0)	4.1 (3.1)	1.540	.063
	WSS Perceptual Aberration Total	2.5 (2.7)	3.0 (3.5)	-.749	.228
	CAARMS Unusual Thinking Content	4.8 (1.9)	5.1 (1.6)	-.651	.258
	CAARMS Disorganized Speech	2.4 (2.2)	1.3 (1.6)	2.497	.007*
	CAARMS Non Bizarre Ideas	5.06 (1.9)	4.6 (2.1)	1.042	.150
	CAARMS Perceptual Abnormalities	3.1 (2.5)	3.4 (2.5)	-.505	.307
Imaging Classification					
	PANSS Positive Total	20.4 (6.3)	18.2 (6.0)	1.495	.069
	WSS Magical Ideation Total	4.2 (2.7)	4.5 (3.2)	-.385	.350
	WSS Perceptual Aberration Total	2.5 (2.9)	2.9 (3.3)	-.605	.258

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

|				
CAARMS Unusual Thinking Content	5.3 (1.2)	4.9 (1.9)	1.073	.143
CAARMS Disorganized Speech	2.2 (1.7)	1.6 (1.9)	1.392	.083
CAARMS Non Bizarre Ideas	4.3 (2.4)	4.9 (1.9)	-1.187	.119
CAARMS Perceptual Abnormalities	3.6 (2.4)	3.2 (2.5)	.770	.215

341

2.3 GMV differences between groups

343   We also wanted to understand the neurobiology that was driving the classifications and

344   misclassifications in our models. We therefore performed whole brain GMV comparisons a)

345   between the whole ROP group and the whole ROD group b) between subjects from all

346   groups who were classified as ROP (regardless of whether it was a correct classification or a

347   misclassification) and c) between Pure subjects and co-morbid subjects. In figure S5 (a,b,c)

348   we can see the GMV differences between the whole of the ROP group and the whole of the

349   ROD group. The ROP group had increased GMV in the superior and inferior frontal gyrus

350   whereas decreased GMV areas were detected in the left cerebellum, the Supramarginal

351   gyrus, and the inferior temporal gyrus when compared to the ROP group (Table S9).

352   Furthermore we compared subjects who were classified as ROP to subject who were

353   classified as ROD according to our imaging model (figure S6 (a,b,c)). Subjects who were

354   classified as ROP showed increased GMV in the insula, the caudate, and the precuneus

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

355   among others, whereas they showed decreased GMV in the lingual gyrus and the thalamus

356   (Table S10). Finally we performed a whole brain GMV analysis between patients who were

357   in Pure groups (both ROP and ROD) and patients who were in co-morbid groups (both

358   ROP+D and ROD+P) (figure s7 (a,b,c)). Pure subjects showed pronounced GMV increases

359   in the left cerebellum and the middle occipital gyrus and decreased GMV in the superior

360   frontal gyrus and the lingual gyrus (Table S11).
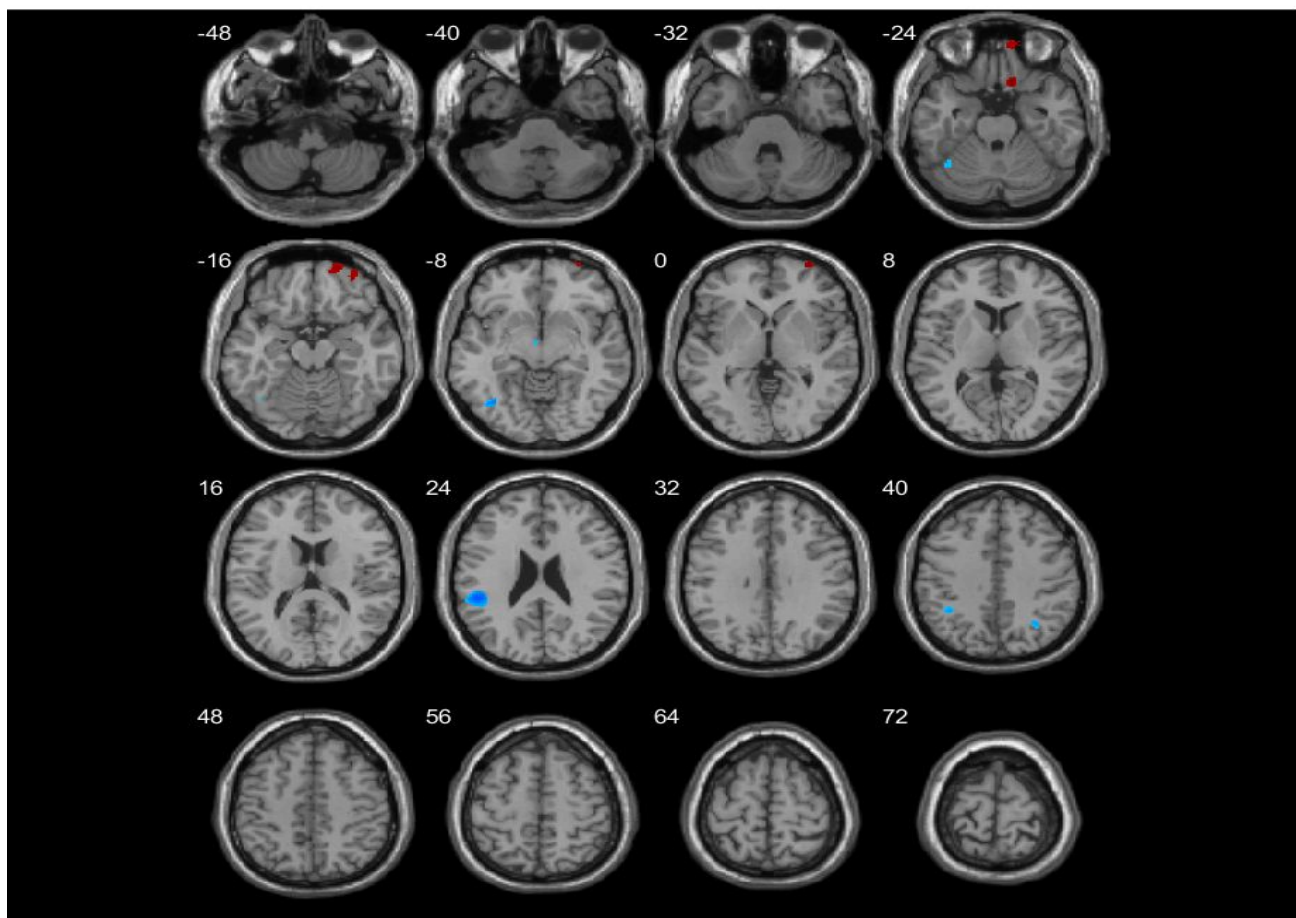
361
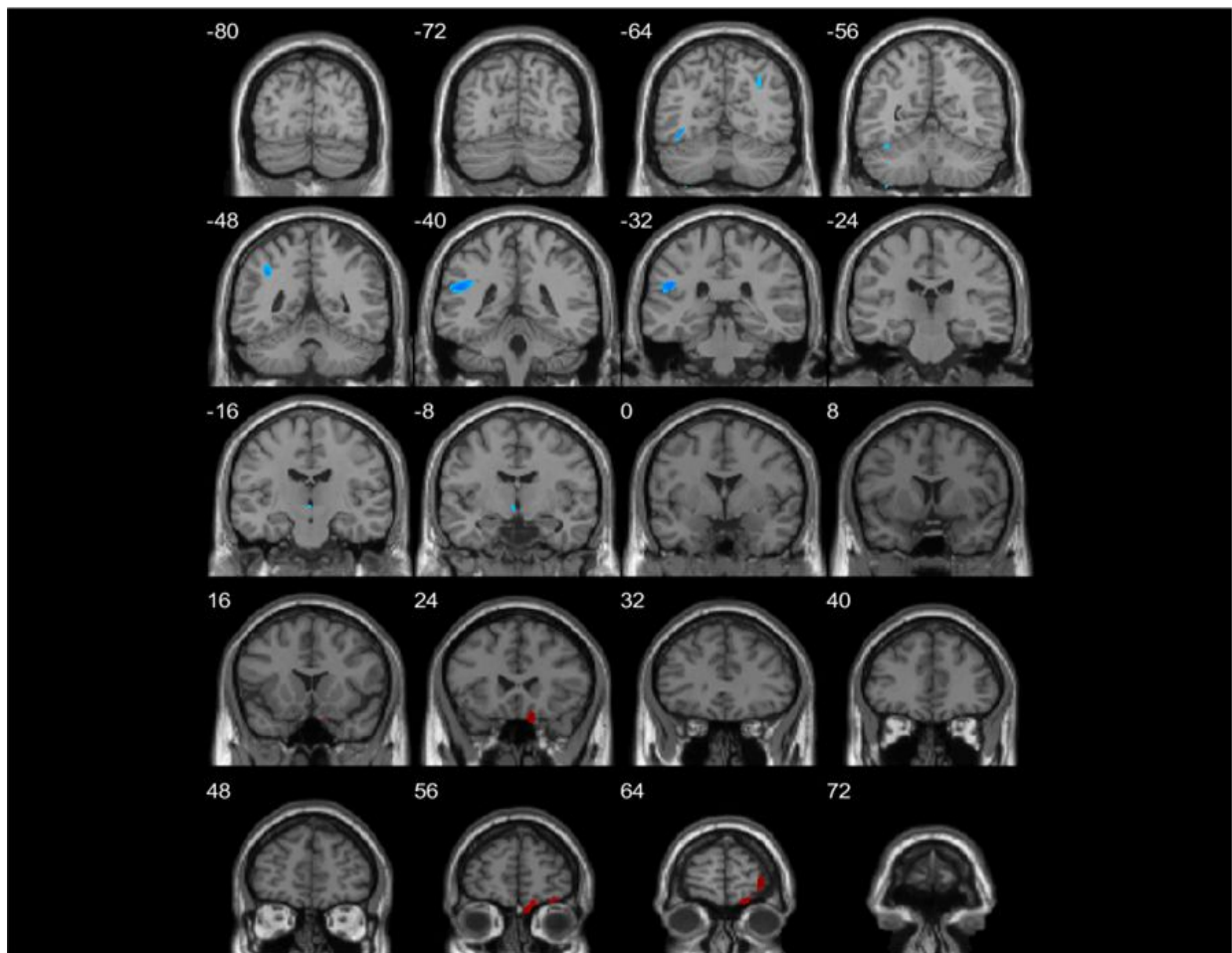
362

363   **Figure S5a. Whole ROP Group vs Whole ROD Brain Differences**

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

364
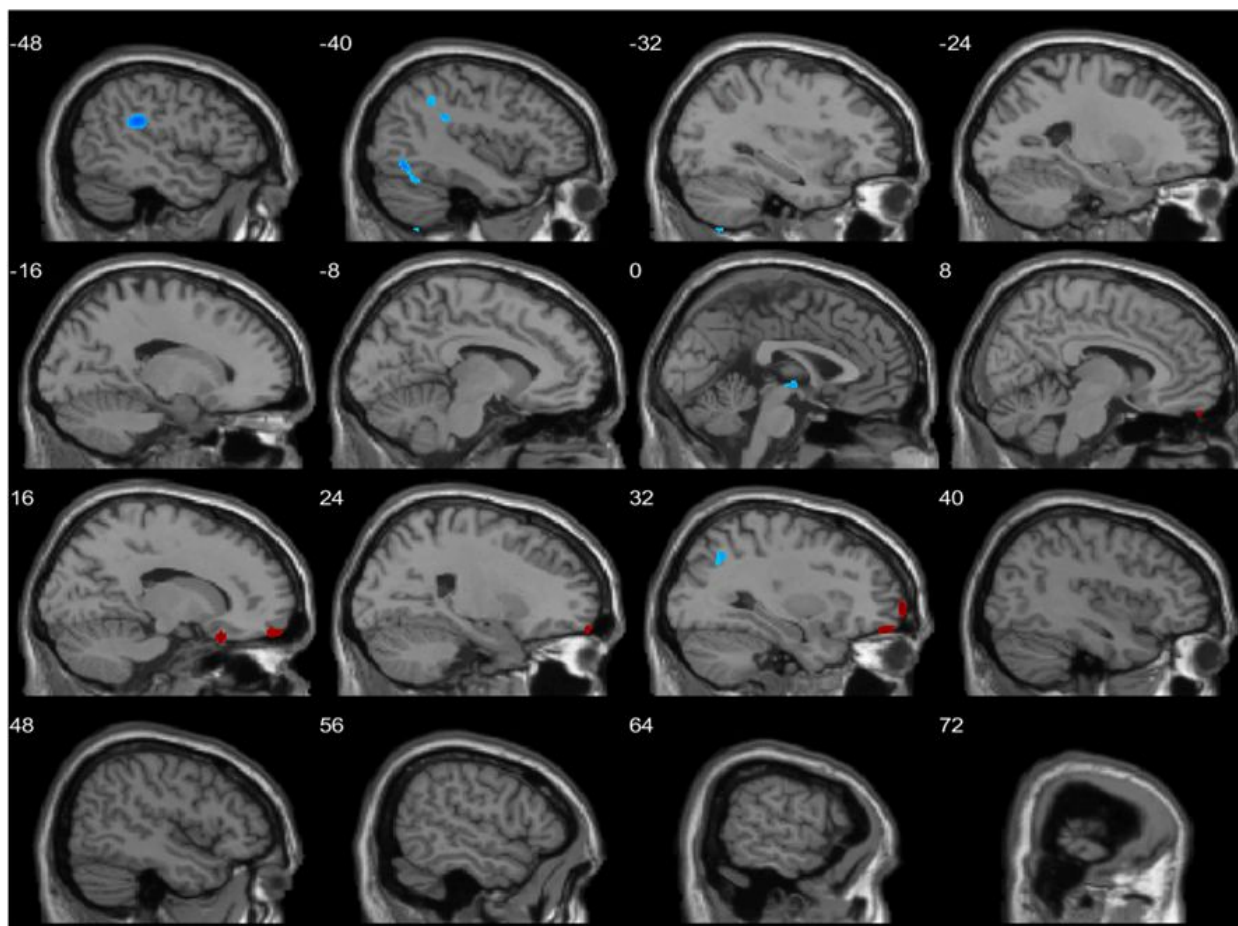
Figure S5b. Whole ROP Group vs Whole ROD Brain Differences

365

366

367

368

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

Figure S5c. Whole ROP Group vs Whole ROD Brain Differences

Lalousis et al. Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

Table S9. Significant clusters from the Whole ROP Group vs Whole ROD Brain Differences analysis. Red=ROP greater GMV than ROD, Blue=ROP less GMV than ROD

Clusters ROP vs ROD

| Brain Area	Cluster Peak Voxel MNI Coordinates	P Value, Peak Intensity, Cluster Size
Superior Frontal Gyrus	20 61 -19	p(FDR)=0.013, T=3.64, k=300
Inferior Frontal Gyrus	14 22 -25	p(FDR)=0.057, T=3.84, k=154
Left Cerebellum	-38 -66 -11	p(FDR)=0.044, T=-4.16, k=208
Supramarginal Gyrus	-50 -36 24	p(FDR)=0.001, T=-4.17, k=548
Inferior Temporal Gyrus	-22 3 -55	p(FDR)=0.001, T=-4.55, k=682

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

381
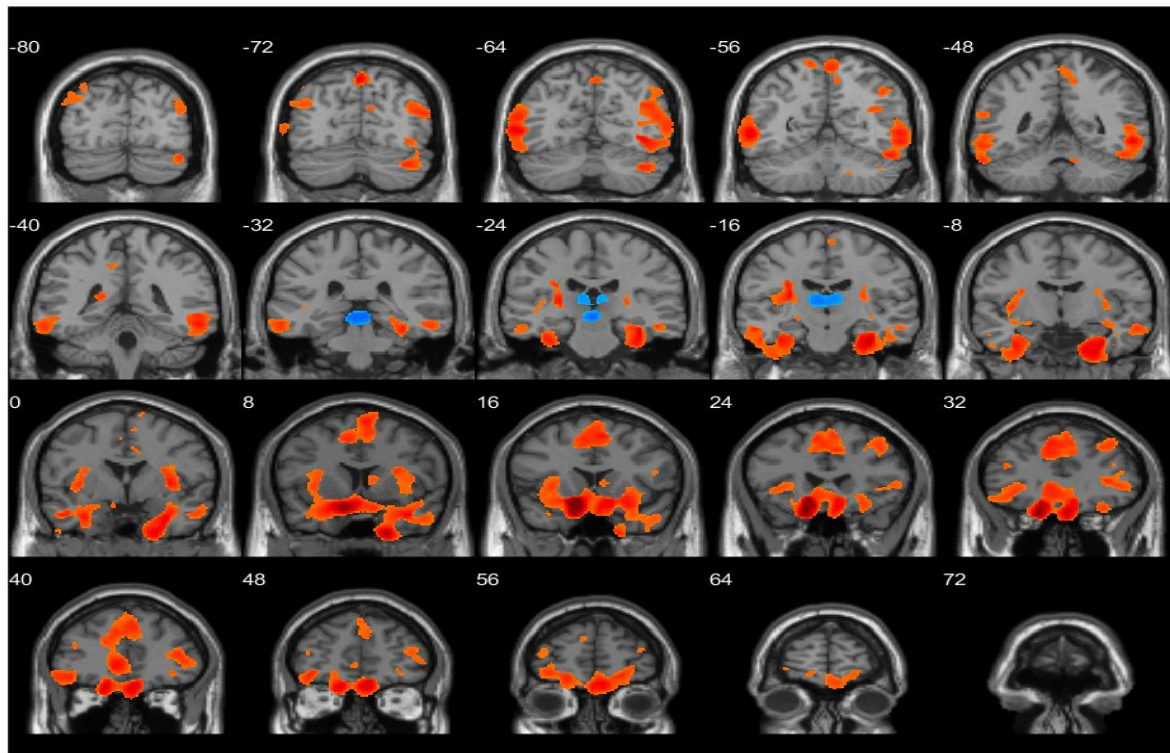
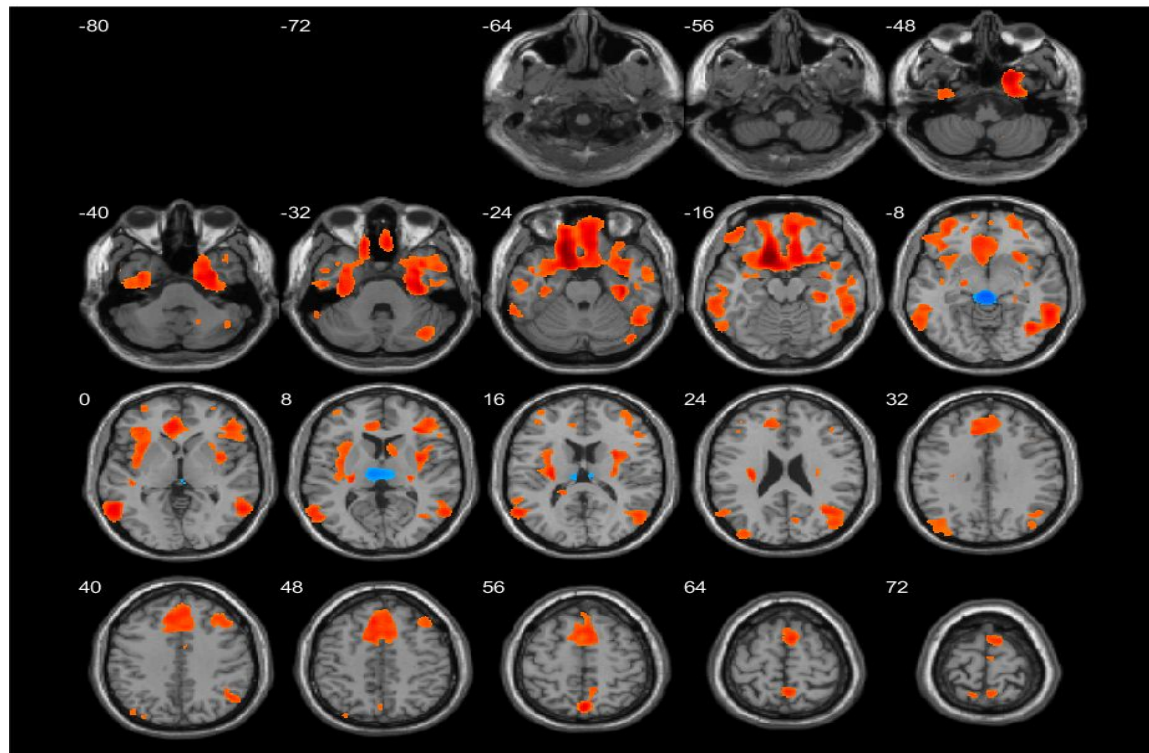**Figure S6a. Subjects Classified as ROP vs Subjects who were classified as ROD**

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

383
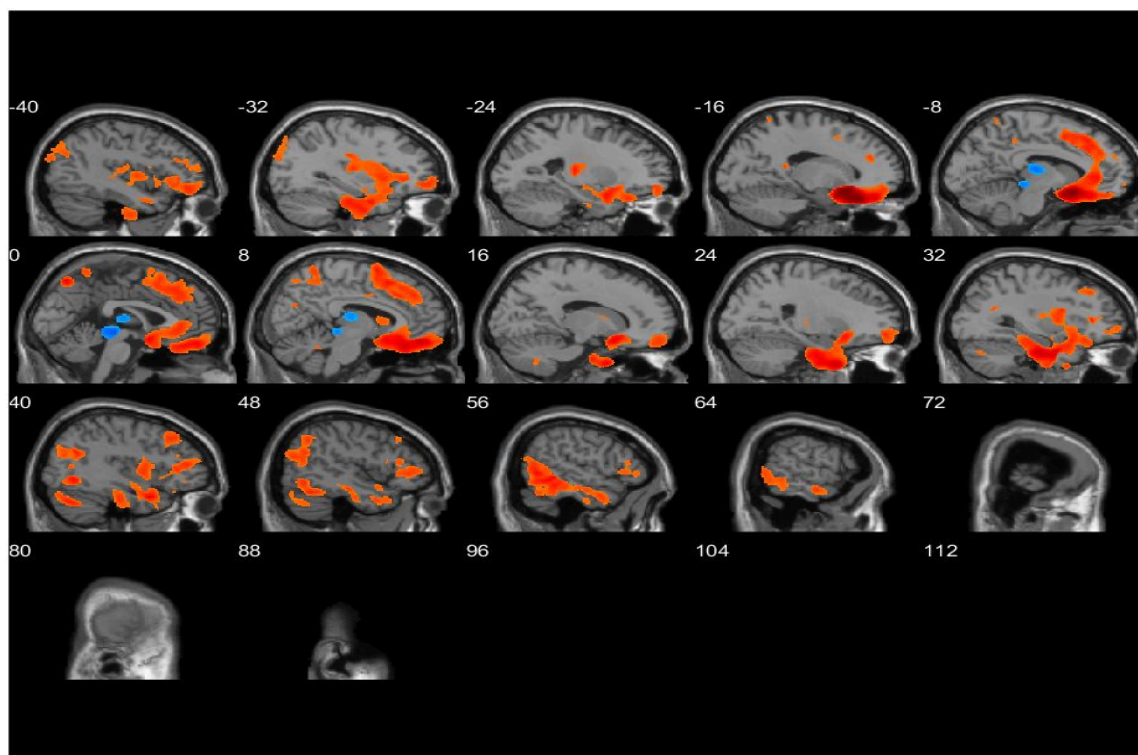
Figure S6b. Subjects Classified as ROP vs Subjects who were classified as ROD

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

385

Figure S6c. Subjects Classified as ROP vs Subjects who were classified as ROD

Table 10. Subjects from all Groups Classified as ROP vs Subjects from all Groups Classified as ROD According to the Imaging Model. Red=ROP greater GMV than ROD, Blue=ROP less GMV than ROD

Clusters Classified as ROP vs Classified as ROD

Brain Area	Cluster Peak Voxel MNI Coordinates	P Value, Peak Intensity, Cluster Size
Middle Temporal Gyrus/Inferior Frontal Gyrus/Insula	-16 18 -13	p(FDR)<0.001, T=6.53, k=61340
Caudate	11 9 6	p(FDR)=0.057, T=3.58, k=288
Middle Frontal Gyrus	-40 34 21	p(FDR)=0.057, T=3.63, k=273
Superior Occipital Gyrus/Precuneus	-35 -86 24	p(FDR)<0.001, T=3.78, k=1117
Middle Frontal Gyrus	40 28 43	p(FDR)<0.001, T=3.98, k=826

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression
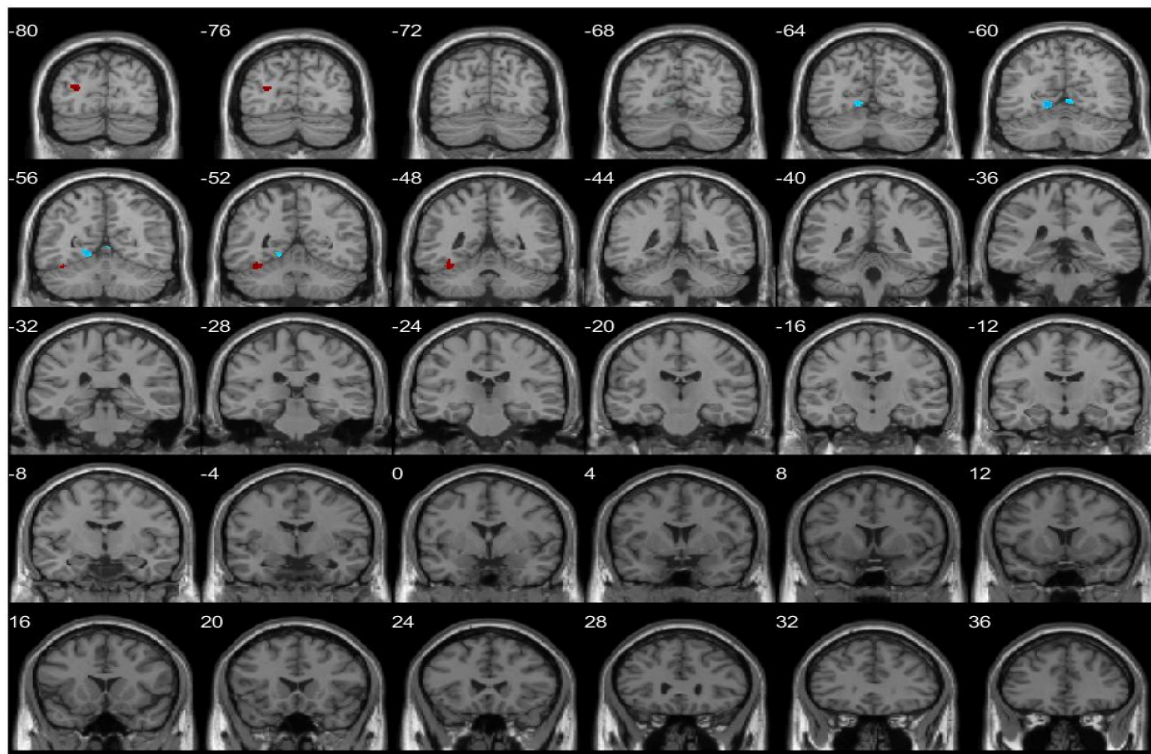
Precuneus	1 -72 55	p(FDR)<0.001, T=5.12, k=1133
Lingual Gyrus	-2 -27 -8	p(FDR)<0.001, T=-4.35, k=888
Thalamus	4 -18 10	p(FDR)<0.001, T=-4.06, k=992

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression
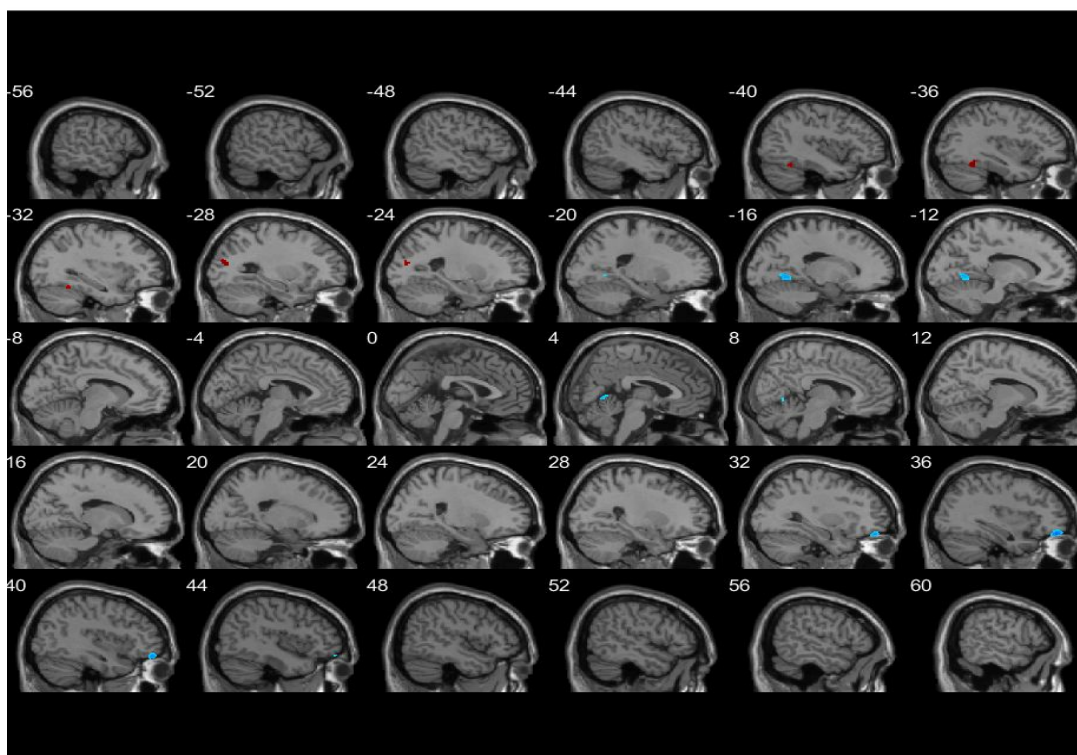
390

391

Figure S7a. Subjects in Pure Groups vs Subjects in Co-Morbid Groups

393
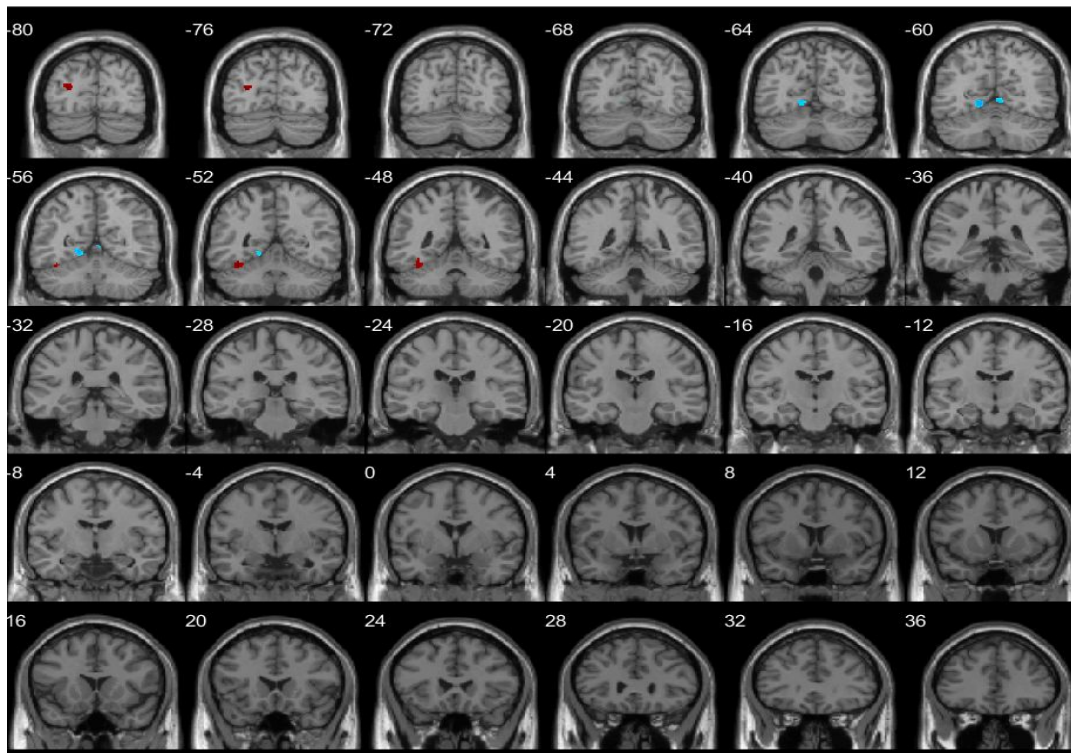
Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

394

Figure S7b. Subjects in Pure Groups vs Subjects in Co-Morbid Groups

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

Figure S7c. Subjects in Pure Groups vs Subjects in Co-Morbid Groups

Table S11. Subjects in Pure Groups (Both ROP and ROD) vs Subjects in Co-Morbid Groups (Both ROP Depressed and ROD Psychotic) Brain Differences. Red=ROP greater GMV than ROD, Blue=ROP less GMV than ROD

Clusters Pure vs Co-Morbid

Brain Area	Cluster Peak Voxel MNI Coordinates	P Value, Peak Intensity, Cluster Size
Left Cerebellum	-38 -51 -21	p(FDR)=0.080, T=3.80, k=155
Middle Occipital Gyrus	-27 -80 17	p(FDR)=0.080, T=3.70, k=120
Superior Frontal Gyrus	37 51 -20	p(FDR)=0.013, T=-4.59, k=304
Lingual Gyrus	-16 -57 -4	p(FDR)=0.018, T=-3.85, k=230

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

401

2.5. Decision Scores Group Differences

403   We wanted to test the differences of the decision scores between the groups (ROP and

404   ROD as well as ROP+D and ROD+P). In the pure clinical model the mean decision score for the

405   ROP group was .4108 (SD=0.805) and for the ROD group -.7760 (SD=0.726). An independent

406   samples t-test showed that the difference between the two groups was significant (t=8.177,

407   df=126, p<.0001). In the pure neuroanatomical model the mean decision score for the ROP group

408   was -.0579 (SD=0.227) and for the ROD group -.2096 (SD=0.178). An independent samples t-

409   test showed that the difference between the two groups was significant (t=4.039, df=126,

410   p=.0001). In the pure combined model the mean decision score for the ROP group was .6988

411   (SD=1.121) and for the ROD group -1.019 (SD=1.012). An independent samples t-test showed

412   that the difference between the two groups was significant (t=8.490, df=126, p<.0001).

413   In the applied clinical model the mean decision score for the ROP group was -.2507

414   (SD=0.698) and for the ROD group -.6742 (SD=0.701). An independent samples t-test showed

415   that the difference between the two groups was significant (t=2.686, df=115, p=.005). In the

416   applied neuroanatomical model the mean decision score for the ROP group was -.1619

417   (SD=1.859) and for the ROD group -.1217 (SD=0.201). An independent samples t-test showed

418   that the difference between the two groups was not significant (t=-.942, df=115, p=.187). Finally,

419   in the applied combined model the mean decision score for the ROP group was -.2798

420   (SD=1.003) and for the ROD group -.8769 (SD=1.008). An independent samples t-test showed

421   that the difference between the two groups was significant (t=2.635, df=115, p=.006).

422

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and   depression

2.6. Decision Scores Regression with 9 Month Functional Outcome

Finally we wanted to test whether decision scores were related to 9 month global functioning outcome using the Global Functioning:

Social (GFS) and Global Functioning Role (GFR) scales. Good functioning was interpreted as a score on either scale of more than 7, and the

impaired functioning was interpreted as score on either scale of 7 or less[46]. A logistic regression was performed to determine the likelihood

that patients had an impaired functional outcome at 9 months using decision scores from our SVM models as predictors. None of the models

was statistically significant. GFS model using pure SVM model decision scores: $\chi 2(3) = 2.259$, $p = .520$; GFR model using pure SVM model

decision scores: $\chi 2(3) = 3.638$, $p = .303$; GFS model using comorbid SVM model decision scores: $\chi 2(3) = 7.179$, $p = .066$; GFR model using

comorbid SVM model decision scores: $\chi 2(3) = 3.353$, $p = .340$. Table S12 contains information from the logistic regression models.

Table S12. Odds Ratios (OR) for Impaired Functional Outcomes for Decision Scores from the SVM models

Global Functioning Scale	Outcome	No.	Beta/OR (95% Confidence Intervals)					
			Pure Clinical & Neurocognitive Model	Pure GMV Model	Pure Stacking Model	Comorbid Clinical & Neurocognitive Model	Comorbid GMV Model	Comorbid Stacking Model
GFR	Impaired	Pure: 70 Comorbid: 84						
	Good	Pure: 58 Comorbid: 33	-1.372/.254 (.008-7.812)	-.956/.384 (.063-2.348)	.814/.2.256 (.200-25.406)	23.189/1178.0 (.003-4263.0)	.319/1.375 (.155-12.175)	-15.961/.000 (.000-61.799)
GFS	Impaired	Pure: 69 Comorbid: 77						
	Good	Pure: 59 Comorbid: 40	1.474/4.365 (.138-137.5)	-1.023/.359 (.060-2.144)	-.912/.402 (.035-4.582)	34.917/1460.0 (251.3-8478.0)	.564/1.758 (.215-14.389)	-24.104/.000 (.000-.025)

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

432 **2.7. Chi-Square Analysis of Misclassifications in the Replication Sample**

433 In the clinical/neurocognitive model, ROP+D patients were more frequently classified as ROD

434 compared to pure ROP patients (χ^2=18.878; p<0.001). In contrast, the assignment precision of

435 ROD+P and ROD patients did not differ (χ^2=.379; p=0.538). Similarly to the

436 clinical/neurocognitive model ROP+D patients were more frequently classified as ROD

437 compared to pure ROP patients (χ^2=.003; p=0.958), at a statistically not significant level

438 potentially driven by the fact that the majority of pure ROP patients were classified as ROD. The

439 assignment precision of ROD+P and ROD patients did not differ (χ^2=0.112; p=0.738).

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

455

#The PRONIA consortium:

The authors listed here performed the screening, recruitment, rating, examination, and follow-up of the study participants. They were involved in implementing the examination protocols of the study, setting up its IT infrastructure, and organizing the flow and quality control of the data analyzed in this manuscript between the local study sites and the central study database.

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

479 **Department of Psychiatry and Psychotherapy, Ludwig-Maximilian-University, Munich,**
480 **Germany**

481 Linda Betz, Anne Erkens, Eva Gussmann, Shalaila Haas, Alkomiet Hasan, Claudius Hoff, Ifrah
482 Khanyaree, Aylin Melo, Susanna Muckenhuber-Sternbauer, Janis Köhler, Ömer Öztürk, Nora
483 Penzel, David Popovic, Adrian Rangnick, Sebastian von Saldern, Rachele Sanfelici, Moritz
484 Spangemacher, Ana Tupac, Maria Fernanda Urquijo, Johanna Weiske, Antonia Wosgien

485 **Department of Psychiatry and Psychotherapy, University of Cologne, Cologne, Germany**

486 Karsten Blume, Dominika Gebhardt, Nathalie Kaiser, Ruth Milz, Alexandra Nikolaides, Mauro
487 Seves, Silke Vent, Martina Wassen

488 **Department of Psychiatry (Psychiatric University Hospital, UPK), University of Basel,**
489 **Switzerland**

490 Christina Andreou, Laura Egloff, Fabienne Harrisberger, Claudia Lenz, Letizia Leanza, Amatya
491 Mackintosh, Renata Smieskova, Erich Studerus, Anna Walter, Sonja Widmayer

492 **Institute for Mental Health & School of Psychology, University of Birmingham, United**
493 **Kingdom**

494 Chris Day, Mariam Iqbal, Mirabel Pelton, Pavan Mallikarjun, Alexandra Stainton, Ashleigh Lin

495 **Department of Psychiatry, University of Turku, Finland**

496 Alexander Denissoff, Anu Ellilä, Tiina From, Markus Heinimaa, Tuula Ilonen, Päivi Jalo, Heikki
497 Laurikainen, Antti Luutonen, Akseli Mäkela, Janina Paju, Henri Pesonen, Reetta-Liina Säilä,
498 Anna Toivonen, Otto Turtonen

499 **General Electric Global Research Inc., USA**

500 Ana Beatriz Solana, Manuela Abraham, Nicolas Hehn, Timo Schirmer

501 **Workgroup of Paolo Brambilla, University of Milan, Italy:**

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

502 **Department of Neuroscience and Mental Health, Fondazione IRCCS Ca' Granda Ospedale**

503 **Maggiore Policlinico, University of Milan, Milan, Italy:** Carlo Altamura, Marika Belleri,

504 Francesca Bottinelli, Adele Ferro, Marta Re

505 **Programma2000, Niguarda Hospital, Milan:** Emiliano Monzani, Maurizio Sberna

506 **San Paolo Hospital, Milan:** Armando D'Agostino, Lorenzo Del Fabro

507 **Villa San Benedetto Menni, Albese con Cassano (CO**): Giampaolo Perna, Maria Nobile,

508 Alessandra Alciati

509 **Workgroup of Paolo Brambilla at the University of Udine, Italy**

510 **Department of Medical Area, University of Udine, Udine, Italy:** Matteo Balestrieri, Carolina

511 Bonivento, Giuseppe Cabras, Franco Fabbro

512 **IRCCS Scientific Institute "E. Medea", Polo FVG, Udine:** Marco Garzitto, Sara Piccin

513

514

515

516

517

518

519

520

521

522

523

524

525

526

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

527

528

529

530

References

1.   Yung AR, Yuen HP, McGorry PD, et al. Mapping the onset of psychosis: the Comprehensive Assessment of At-Risk Mental States. *Aust N Z J Psychiatry*. 2005;39(11-12):964-971. doi:10.1080/j.1440-1614.2005.01714.x

2.   Llorca P-M, Lançon C, Lancrenon S, et al. The "Functional Remission of General Schizophrenia" (FROGS) scale: Development and validation of a new questionnaire. *Schizophrenia Research*. 2009;113(2-3):218-225. doi:10.1016/j.schres.2009.04.029

3.   Cornblatt BA, Auther AM, Niendam T, et al. Preliminary findings for two new measures of social and role functioning in the prodromal phase of schizophrenia. *Schizophr Bull*. 2007;33(3):688-702. doi:10.1093/schbul/sbm029

4.   Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull*. 1987;13(2):261-276.

5.   Alvarez E, Garcia-Ribera C, Torrens M, Udina C, Guillamat R, Casas M. Premorbid adjustment scale as a prognostic predictor for schizophrenia. *Br J Psychiatry*. 1987;150:411. doi:10.1192/bjp.150.3.411

6.   Andreasen NC. The Scale for the Assessment of Negative Symptoms (SANS): conceptual and theoretical foundations. *Br J Psychiatry Suppl*. 1989;(7):49-58.

7.   First, Michael B, Spitzer, Robert L, Gibbon, Miriam, Williams, Janet B.W. *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition With Psychotic Screen (SCID-I/P W/ PSY SCREEN)*. Biometrics Research, New York State Psychiatric Institute.; 2002.

8.   McGlashan TH, Miller TJ, Woods SW, Hoffman RE, Davidson L. Instrument for the Assessment of Prodromal Symptoms and States. In: Miller T, Mednick SA, McGlashan TH, Libiger J, Johannessen JO, eds. *Early Intervention in Psychotic Disorders*. NATO Science Series. Springer Netherlands; 2001:135-149. doi:10.1007/978-94-010-0892-1_7

9.   F. Schultze-Lutter, J. Addington, S. Ruhrmann, J. Klosterkötter. *Schizophrenia Proneness Instrument, Adult Version (SPI-A)*. Giovanni Fioriti Editore.; 2007.

10. Beck A, Steer R, Brown G. Manual for the Beck Depression Inventory-II (BDI-II). Published online January 1, 1996. Accessed February 24, 2020. https://www.scienceopen.com/document?vid=9feb932d-1f91-4ff9-9d27-da3bda716129

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

561   11.   Endler NS. *CISS: Coping Inventory for Stressful Situations*. Harcourt; 2004.

562   12.   Bernstein DP, Fink L, Handelsman L, et al. Initial reliability and validity of a new
563   retrospective measure of child abuse and neglect. *Am J Psychiatry*. 1994;151(8):1132-1136.
564   doi:10.1176/ajp.151.8.1132

565   13.   Veale JF. Edinburgh Handedness Inventory—Short Form: A revised version based on
566   confirmatory factor analysis. *Laterality: Asymmetries of Body, Brain and Cognition*.
567   2014;19(2):164-177. doi:10.1080/1357650X.2013.783045

568   14.   Williams DR. Measuring Discrimination Resource. :24.

569   15.   Cole JD, Kazarian SS. The Level of Expressed Emotion Scale: A new measure of expressed
570   emotion. *Journal of Clinical Psychology*. 1988;44(3):392-397. doi:10.1002/1097-
571   4679(198805)44:3<392::AID-JCLP2270440313>3.0.CO;2-3

572   16.   Zimet GD, Dahlem NW, Zimet SG, Farley GK. The Multidimensional Scale of Perceived
573   Social Support. *Journal of Personality Assessment*. 1988;52(1):30-41.
574   doi:10.1207/s15327752jpa5201_2

575   17.   Costa PT, McCrae RR, Psychological Assessment Resources I. *Revised NEO Personality
576   Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI)*. Psychological
577   Assessment Resources; 1992.

578   18.   Friborg O, Hjemdal O, Rosenvinge JH, Martinussen M. A new rating scale for adult
579   resilience: what are the central protective resources behind healthy adjustment? *Int J
580   Methods Psychiatr Res*. 2003;12(2):65-76.

581   19.   Connor KM, Davidson JR, Churchill LE, Sherwood A, Foa E, Weisler RH. Psychometric
582   properties of the Social Phobia Inventory (SPIN). New self-rating scale. *Br J Psychiatry*.
583   2000;176:379-386. doi:10.1192/bjp.176.4.379

584   20.   Health WHOD of M. WHOQOL-BREF : introduction, administration, scoring and generic
585   version of the assessment : field trial version, December 1996. Published online 1996.
586   Accessed April 6, 2020. https://apps.who.int/iris/handle/10665/63529

587   21.   Cornblatt BA, Risch NJ, Faris G, Friedman D, Erlenmeyer-Kimling L. The Continuous
588   Performance Test, identical pairs version (CPT-IP): I. New findings about sustained
589   attention in normal families. *Psychiatry Res*. 1988;26(2):223-238. doi:10.1016/0165-
590   1781(88)90076-5

591   22.   Nowicki S, Duke MP. Nonverbal receptivity: The Diagnostic Analysis of Nonverbal
592   Accuracy (DANVA). In: *Interpersonal Sensitivity: Theory and Measurement*. The LEA
593   series in personality and clinical psychology. Lawrence Erlbaum Associates Publishers;
594   2001:183-198.

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

595   23.   Rey A. L'examen psychologique dans les cas d'encéphalopathie traumatique. (Les
596   problems.). [The psychological examination in cases of traumatic encepholopathy.
597   Problems.]. *Archives de Psychologie*. 1941;28:215-285.

598   24.   Roiser JP, Stephan KE, den Ouden HEM, Friston KJ, Joyce EM. Adaptive and aberrant
599   reward prediction signals in the human brain. *Neuroimage*. 2010;50(2):657-664.
600   doi:10.1016/j.neuroimage.2009.11.075

601   25.   Petrides M, Milner B. Deficits on subject-ordered tasks after frontal- and temporal-lobe
602   lesions in man. *Neuropsychologia*. 1982;20(3):249-262. doi:10.1016/0028-3932(82)90100-2

603   26.   Reitan RM. Validity of the Trail Making Test as an Indicator of Organic Brain Damage:
604   *Perceptual and Motor Skills*. 1958;8(3). doi:10.2466/pms.1958.8.3.271

605   27.   Manjón JV, Tohka J, García-Martí G, et al. Robust MRI brain tissue parameter estimation
606   by multistage outlier rejection. *Magnetic Resonance in Medicine*. 2008;59(4):866-873.
607   doi:10.1002/mrm.21521

608   28.   Rajapakse JC, Giedd JN, Rapoport JL. Statistical approach to segmentation of single-
609   channel cerebral MR images. *IEEE Trans Med Imaging*. 1997;16(2):176-186.
610   doi:10.1109/42.563663

611   29.   Mushquash C, O'Connor BP. SPSS and SAS programs for generalizability theory analyses.
612   *Behavior Research Methods*. 2006;38(3):542-547. doi:10.3758/BF03192810

613   30.   Wolpert DH. Stacked generalization. *Neural Networks*. 1992;5(2):241-259.
614   doi:10.1016/S0893-6080(05)80023-1

615   31.   R. Polikar. Ensemble based systems in decision making - IEEE Journals & Magazine.
616   Published 2006. Accessed May 20, 2019. https://ieeexplore.ieee.org/document/1688199

617   32.   Saeys Y, Inza I, Larrañaga P. A review of feature selection techniques in bioinformatics.
618   *Bioinformatics*. 2007;23(19):2507-2517. doi:10.1093/bioinformatics/btm344

619   33.   Fan R-E, Chang K-W, Hsieh C-J, Wang X-R, Lin C-J. LIBLINEAR: A Library for Large
620   Linear Classification. *J Mach Learn Res*. 2008;9:1871–1874.

621   34.   Cannon-Spoor HE, Potkin SG, Wyatt RJ. Measurement of Premorbid Adjustment in
622   Chronic Schizophrenia. *Schizophr Bull*. 1982;8(3):470-484. doi:10.1093/schbul/8.3.470

623   35.   Barkus E, Badcock JC. A Transdiagnostic Perspective on Social Anhedonia. *Front
624   Psychiatry*. 2019;10:216. doi:10.3389/fpsyt.2019.00216

625   36.   Shankman SA, Nelson BD, Harrow M, Faull R. Does physical anhedonia play a role in
626   depression? A 20-year longitudinal study. *J Affect Disord*. 2010;120(1-3):170-176.
627   doi:10.1016/j.jad.2009.05.002

Lalousis et al.   Diagnosis, Comorbidity, and Classification in recent onset psychosis and depression

628   37.   Romm KL, Rossberg JI, Hansen CF, Haug E, Andreassen OA, Melle I. Self-esteem is
629   associated with premorbid adjustment and positive psychotic symptoms in early psychosis.
630   *BMC Psychiatry*. 2011;11(1):136. doi:10.1186/1471-244X-11-136

631   38.   Silverstone PH, Salsali M. Low self-esteem and psychiatric patients: Part I – The
632   relationship between low self-esteem and psychiatric diagnosis. *Annals of General Hospital*
633   *Psychiatry*. 2003;2(1):2. doi:10.1186/1475-2832-2-2

634   39.   Kline ER, Seidman LJ, Cornblatt BA, et al. Depression and clinical high-risk states:
635   Baseline presentation of depressed vs. non-depressed participants in the NAPLS-2 cohort.
636   *Schizophr Res*. 2018;192:357-363. doi:10.1016/j.schres.2017.05.032

637   40.   Schultze-Lutter F, Ruhrmann S, Picker H, von Reventlow HG, Brockhaus-Dumke A,
638   Klosterkötter J. Basic symptoms in early psychotic and depressive disorders. *Br J*
639   *Psychiatry Suppl*. 2007;51:s31-37. doi:10.1192/bjp.191.51.s31

640   41.   Han M-H, Nestler EJ. Neural Substrates of Depression and Resilience. *Neurotherapeutics*.
641   2017;14(3):677-686. doi:10.1007/s13311-017-0527-x

642   42.   Marulanda S, Addington J. Resilience in Individuals at Clinical High Risk for Psychosis.
643   *Early Interv Psychiatry*. 2016;10(3):212-219. doi:10.1111/eip.12174

644   43.   McClintock SM, Husain MM, Greer TL, Cullum CM. Association between depression
645   severity and neurocognitive function in major depressive disorder: A review and synthesis.
646   *Neuropsychology*. 2010;24(1):9-34. doi:10.1037/a0017336

647   44.   Sánchez-Torres AM, Elosúa MR, Lorente-Omeñaca R, Moreno-Izco L, Cuesta MJ. A
648   comparative study of the working memory multicomponent model in psychosis and healthy
649   controls. *Comprehensive Psychiatry*. 2015;61:97-105.
650   doi:10.1016/j.comppsych.2015.05.008

651   45.   Becker HE, Nieman DH, Dingemans PM, van de Fliert JR, De Haan L, Linszen DH. Verbal
652   fluency as a possible predictor for psychosis. *Eur Psychiatry*. 2010;25(2):105-110.
653   doi:10.1016/j.eurpsy.2009.08.003

654   46.   Koutsouleris N, Kambeitz-Ilankovic L, Ruhrmann S, et al. Prediction Models of Functional
655   Outcomes for Individuals in the Clinical High-Risk State for Psychosis or With Recent-
656   Onset Depression: A Multimodal, Multisite Machine Learning Analysis. *JAMA Psychiatry*.
657   2018;75(11):1156-1172. doi:10.1001/jamapsychiatry.2018.2165

658