

Title: Is there a symptomatic distinction between the affective psychoses and schizophrenia? A machine learning approach

S Jauhar¹, R Krishnadas², M M Nour¹, D Cunningham-Owens³, E C Johnstone³, S M Lawrie³

¹Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, Denmark Hill, London, SE5 8AF.

²Sackler Institute for Psychobiological Research, Institute of Neurological Sciences, Southern General Hospital, Glasgow G51 4TF.

³Division of Psychiatry, Kennedy Tower, Royal Edinburgh Hospital, Edinburgh EH10 5HF.

Corresponding Author- Sameer Jauhar

Telephone: +44 7957 335 722

Fax +44 141 427 6661

Email Sameer.jauhar@kcl.ac.uk

Address PO 63, Department of Psychosis Studies, King's College, London SE5 8AF

Email addresses of other authors: Rajeev.krishnadas@glasgow.ac.uk ,
Matthew.Nour@kcl.ac.uk , David.Owens@ed.ac.uk , E.Johnstone@ed.ac.uk ,
S.Lawrie@ed.ac.uk

Abstract

Dubiety exists over whether clinical symptoms of schizophrenia can be distinguished from affective psychosis, the assumption being that absence of a “point of rarity” indicates lack of nosological distinction, based on prior group-level analyses. Using psychopathology ratings from people with “functional psychosis”, assigned DSM III diagnoses two and a half years later, we examined whether initial clinical syndromes could subsequently distinguish diagnostic constructs at an individual level. Advanced machine learning techniques, using unsupervised (hierarchical clustering) and supervised (regularized logistic regression algorithm and nested-cross-validation) were applied to a dataset of 214 patients with functional psychosis (schizophrenia n=120, affective psychosis, n=82). Patients were initially assessed with the Present State Examination (PSE), and followed up 2.5 years later, when DSM III diagnoses were applied (independent of initial PSE).

Based on PSE syndromes, unsupervised learning discriminated depressive and mania/psychosis clusters (approximately unbiased probability, AUP, 0.92 and 0.94), which split into two groups (manic and psychosis) after removal of the depressive group (AUP 0.84 and 0.88). Supervised machine learning classified schizophrenia or affective psychosis with 83.66% (95% CI = 77.83% to 88.48%) accuracy. Area under the ROC curve (AUROC) was 89.14 %. True positive rate for schizophrenia was 88.24% (95%CI = 81.05 - 93.42%) and affective psychosis 77.11% (95%CI = 66.58 - 85.62). Classification accuracy and AUROC remained high when PSE syndromes corresponding to affective symptoms (i.e. used to distinguish DSM III diagnosis) were removed. PSE syndromes, based on clinical symptoms, therefore discriminated between schizophrenia and affective psychosis, suggesting validity to

these diagnostic constructs.

Keywords; Classification, schizophrenia, psychosis, bipolar disorder, psychopathology, nosology, first rank symptoms, machine learning.

(Word count; 247 words)

Introduction

Over the last hundred years arguments have been put forward for, and against, the concept of schizophrenia as a valid construct. Kraepelin's conceptualised schizophrenia as "dementia praecox", which he suggested could be differentiated from manic depressive illness on the basis of both clinical picture and outcome (Kraepelin, 1987). This dichotomous view of psychotic illness has been the cornerstone of modern psychiatric classification systems, including the recent DSM V (American Psychiatric Association, 2013; World Health Organization, 1993). This continues to have clinical relevance, treatments such as lithium having efficacy in affective psychoses, in comparison to schizophrenia (Lawrie et al., 2010). This has recently been challenged, with comment made that there is no validity to the diagnostic constructs (Guloksuz and Van Os, 2018).

Evidence has been put forward to suggest that a distinction between the "functional" psychoses cannot be made, either on clinical grounds (Linscott et al., 2010b) or on the basis of advances in molecular genetics (Craddock and Owen, 2007). Essentially, the case has been made that psychotic disorder itself is a continuum, with no sharp demarcation between affective psychoses and schizophrenia in terms of etiopathogenesis and psychopathology. It has been stated that "formal studies of symptom profiles... have typically failed to find a clear discontinuity between the clinical features of the two categories" (Jablensky, 2010), this lack of distinction acknowledged as a "fact" within the schizophrenia literature, worthy of further study (Tandon et al., 2008). The initial assertions draw largely from early discriminant function analysis by Kendell and Gourlay (Kendell and Gourlay, 1970), in which they were unable to separate a

cohort of 292 patients with functional psychoses into schizophrenia or affective psychosis on the basis of clinical data, and were unable to show symptomatic “point of rarity”. Subsequent analysis showed a discrimination, after addition of functional outcome (Brockington et al., 1979), subsequent analysis of some of this cohort finding a similar distinction, based on symptom profiles (Kendell et al., 1979). Pertinently, the multivariate techniques used when assessing psychopathology in psychosis cohorts (discriminant function analysis (Brockington et al., 1979; Kendell and Gourlay, 1970), latent class analysis (Murray et al., 2005), grade of membership analysis (Manton et al., 1994; Pomarol-Clotet et al., 2010) and cluster analysis (Jablensky et al., 1993)) classify at a group, as opposed to individual level.

Machine learning is the process by which a computer programme learns how to execute a task, without explicit instruction, by using data. Originally described in 1959, when relating pattern recognition and the game of checkers (Samuel, 1959), examples include email filtering and optical character recognition (OCR) software. These techniques are increasingly used with high-dimensional data (e.g. neuroimaging data), to predict group membership or prognosis.

Recently, similar analyses have been used with symptom data, to predict transition to psychosis in people within a high risk population, though we are unaware of this being used in people with established psychosis (Mechelli et al., 2016). Advanced machine learning is used to ascertain a model that represents the potential relationship between predictors (e.g. symptoms) and class membership (e.g. diagnosis) in a “training” subset. The model is then tested in a “test” subset (independent of the one with which the model was built, the “training dataset”) to

examine if the relationship discovered in the first step is generalizable. A strength of these techniques is the ability to examine model accuracy to classify at an individual, as opposed to group level.

Given the reliance on clinical interview, and continued debates in the literature where the validity of current diagnostic systems (differentiating psychotic illness on the basis of significant mood disorder) has been questioned (Guloksuz and Van Os, 2018), we wished to revisit the question of whether clinical symptoms could differentiate schizophrenia from affective psychoses with reasonable accuracy, at an individual level, using machine learning techniques. Utilizing the historical Northwick Park "functional psychosis" trial dataset, we examined if PSE syndromes collected independent of DSM III diagnosis would group into meaningful clusters and differentiate schizophrenia (non-affective) from affective psychosis with reasonable accuracy. We hypothesized that hierarchical clustering would group PSE syndromes into meaningful clusters, and supervised machine learning (regularized logistic regression) would differentiate schizophrenia (non-affective) from affective psychosis (depression / mania with psychotic features) with reasonable accuracy.

MATERIAL AND METHODS

Ethical standards

The study was conducted under the auspices and according to the rules of the Ethical Committee of the Harrow Health District.

Dataset

We analyzed the Northwick Park “functional psychosis” trial dataset (E.C. Johnstone et al., 1992). These patients were recruited from 360 admissions of 326 individual patients referred to Northwick Park Hospital, Middlesex, with definite or possible psychosis, aged 16 to 69, between August 1982 and October 1986. All of the affective group were experiencing psychotic symptoms. Those with repeated admissions were excluded from the analysis. Further details of the sample are given in the attached flowchart (Figure 1).

(Insert Figure 1 about here)

Briefly, Present State Examination, Version 9 (PSE) (Wing et al., 1974), was administered to all patients at time of initial presentation, by two raters (ECJ and DCGO). The PSE is a clinician-administered diagnostic interview schedule, measuring the presence of a wide range of psychiatric symptoms, on an ordinal scale of increasing severity, from 0 to 2. For the current analysis, all the PSE scores obtained initially were converted to the syndrome scores by an investigator, who was independent of the initial study (SJ). Further details of PSE syndromes are given in Supplementary Material.

A DSM III diagnosis of schizophrenia (non-affective) vs. affective psychosis was given 2.5 years after the initial PSE administration, based on the patient's clinical notes (ECJ and DCGO). These diagnoses were given independent of the initial PSE administration.

Statistical analysis

Statistical analysis of the data was completed in two steps. Firstly, an unsupervised learning algorithm (hierarchical clustering) was employed on the PSE

syndromes alone to see if variables clustered into meaningful entities (psychotic and affective syndromes). In the second step, we used a supervised learning algorithm (regularized logistic regression) to see if a model could be trained to learn the mapping function from PSE syndromes to the DSM III diagnosis, to see how well PSE syndromes could classify a person (unseen by the training set) as having DSM III schizophrenia or affective psychosis.

Unsupervised learning

To see if the 31 PSE syndromes alone grouped into meaningful clusters, we performed hierarchical clustering, using the package `pvclust` in `r` (Suzuki and Shimodaira, 2006). The package `pvclust` performs hierarchical cluster analysis using an agglomerative algorithm, via the function `hclust` (measure of dissimilarity computed using the correlation function) using the ward's minimum variance method. For each cluster identified by the algorithm, `pvclust` provides Approximately Unbiased Probability (AUP-values), ranging from 0 and 1, computed by multiscale bootstrap resampling, which indicates how strong the cluster is supported by data (larger values indicate stronger support).

Supervised learning

To see if a model could be trained to differentiate DSM III schizophrenia from affective psychosis using the PSE syndromes, we used regularized logistic regression, using the `cvglmnet` function in `mlr` package in `R` (Bischof et al., 2016; Friedman et al., 2010). This was done twice. First using all the 31 PSE syndromes, and then after removing the affective syndromes identified by the clustering analysis. We used this technique for several reasons. Firstly, the PSE syndromes showed multicollinearity. Secondly, we had 31 predictors and 202 subjects and conventional regression methods would have resulted in over-fitting. The

elasticnet penalty implemented in GLMnet alleviates these issues, using regularization. GLMnet fits a generalized linear model via penalized maximum likelihood. The regularization path is computed for the lasso (L1 norm), ridge (L2 norm) or elasticnet (a mix of L1 and L2 norms) penalty at a grid of values for the regularization parameter lambda (λ).

The objective function for the penalized logistic regression uses the negative binomial log-likelihood and is

$$\min_{(\beta_0, \beta) \in \mathbb{R}^{p+1}} - \left[\frac{1}{N} \sum_{i=0}^N y_i \cdot (y_i, \beta_0 + x_i^T \beta) - \log(1 + e^{(\beta_0 + x_i^T \beta)}) \right] + \lambda \left[(1-\alpha) \|\beta\|_2^2 / 2 + \alpha \|\beta\|_1 \right]$$

The algorithm uses a quadratic approximation to the log-likelihood, and then coordinate descent on the resulting penalized weighted least-squares problem.

The *elastic-net* penalty is controlled by α , and bridges the gap between lasso ($\alpha=1$, the default) and ridge ($\alpha=0$). The parameter λ controls the overall strength of the penalty (Friedman et al., 2010). The algorithm was implemented using nested cross validation in mlr (Bischl et al., 2016). Briefly, the outer resampling (10 fold) loop consisted of 10 pairs of training/test sets. The training set is further partitioned into 10 subsets of equal size. For different values of the hyperparameters, the error rate is estimated based on the training set, within the 10-fold cross-validation scheme (inner resampling loop). The hyperparameter values yielding the smallest cross-validated error rate (from the inner loop) are then used for construction of the logistic regression classifier, that is then fitted on each outer training set, and its performance is evaluated on the outer test set. It should be noted that the test dataset is not used for hyperparameter tuning. This method provides a measure of the ability of the classifier to correctly classify

‘unseen’ cases (ie. cases not used for training), and is a measure of the generalizability of the classifier. The classification accuracy or the proportion of samples classified correctly as non-affective psychosis or affective psychosis in the test dataset was noted. In addition to the classification accuracy, we also report the area under the receiver operator characteristic curve (AUC), true positive rates and the precision (positive predictive values, PPV) for both conditions. Here, true positive rate of a given condition A, measures the proportion of those with condition A that are correctly identified by the test. Precision represents the proportion of individuals with the diagnosis A among all those who have tested positive for the condition. Confidence intervals were calculated using the Clopper-Pearson method or the standard logit method (Mercaldo et al., 2007).

Addressing possible circularity of analyses

To avoid circularity, we repeated the above analysis after excluding the PSE syndromes that were identified as affective syndromes by the hierarchical clustering (ideas of reference, non-specific depression, social unease, depersonalisation, obsessional neurosis, hypochondriasis, depressive delusions, worrying, loss of interest, lack of energy, other depressive symptoms, grandiose delusions, irritability, agitation, hypomania and overactivity).

Results:

Demographics

Of the 326 patients (161 males and 165 females) considered for the original study, clinical examination using the PSE was conducted on 318 subjects. After exclusion of those diagnosed with first-episode, organic and “other psychoses”, and cases without adequate information from interview, 214 cases were available for analysis (see Flowchart, Figure 1). Further demographic details of this sample have

been reported elsewhere (12). We also excluded 12 patients with schizoaffective disorder.

Unsupervised learning

Figure 2a and 2b show results of hierarchical clustering. Grossly, each leaf corresponds to a PSE syndrome, and as one moves up the tree, observations that are similar to each other are combined into branches, which are themselves fused at a higher height. The height of the fusion indicates dissimilarity between two observations. The higher the height, the less similar the observations. In figure 2a, the two rectangle boxes suggest that grossly the data groups itself into two clusters, with an AUP of 0.94 (manic/psychotic cluster) and 0.92 (depressive cluster). We then repeated the hierarchical clustering, after removing syndromes in the depressive cluster (Figure 2b). This analysis once again revealed two main clusters, one with an AUP of 0.88 suggestive of a psychosis cluster, and the other with an AUP of 0.84 with features suggestive of a mania cluster.

(Insert Figure 2 about here)

Supervised learning

Using all PSE syndromes

PSE syndrome differentiated DSM III non-affective (schizophrenia) from affective psychosis (mania/depression with psychosis) with 83.66% (95% CI = 77.83% to 88.48%) accuracy. The area under the ROC curve was 89.14 % (Figure 3a).

The percentage of non-affective psychosis correctly identified (true positive rate for non-affective psychosis) was 88.24% (95%CI = 81.05 - 93.42%), the percentage of affective psychosis correctly identified (true positive rate for affective psychosis) was 77.11% (95%CI = 66.58 - 85.62%). The proportion of individuals with non-affective psychosis among those with PSE syndromes contributing to non-

affective psychosis (precision) was 84.68% (95%CI = 78.74% to 89.18%). The proportion of affective psychosis among those with PSE syndromes identified as contributing to affective psychosis (precision) was 82.05 % (95%CI = 73.38% to 89.28%). The likelihood that a given set of syndromes would be expected in a patient with non-affective psychosis compared to the likelihood that that set of syndromes would be expected in a patient with affective psychosis (likelihood ratio +ve) was 3.85 (95%CI = 2.58to 5.75). In other words, the set of syndromes that predicted non-affective psychosis were almost 4 times more likely in non-affective psychosis compared to affective psychosis. Conversely, the set of syndromes that predicted affective-psychosis were 6 times more likely in affective psychosis compared to non-affective psychosis (likelihood ratio +ve = 6.55; 95%CI = 3.95 to 10.84).

Using non-affective PSE syndromes

To avoid circularity, the above analysis was repeated after excluding affective syndromes identified by hierarchical clustering. The classification accuracy remained above chance level at 74.75% (95%CI = 68.18% to 80.59%), and an area under the ROC of 85.9% (Figure 3b). True positive rate for non-affective psychosis remained high at 83.19% (75.24% to 89.42%), however, the true positive rate for affective psychosis dropped to 62.65% (95% CI = 51.34 - 73.03%), suggesting syndromes that contributed to either diagnosis were distinct. The precision for both conditions was above chance 76.15 (95% CI = 70.50% to 81.02%) and 72.22 % (95% CI = 62.78% to 80.03%) for non-affective and affective psychosis respectively. Odds ratios from a full logistic regression model of variables that predicted a diagnosis of non-affective psychosis are shown in Table 2.

(Insert Figure 3 about here)

(Insert Table 2 about here)

Discussion

In a large cohort of people with functional psychoses, we have shown that clinical psychopathology syndromes differentiated affective psychosis from schizophrenia with reasonable accuracy, using unsupervised and supervised machine learning techniques. Firstly, using unsupervised machine learning (hierarchical cluster analysis) we found that affective clusters can be distinguished from non-affective psychosis, and that non-affective syndromes themselves differentiate the two constructs. We then demonstrated reasonable classification accuracy (>80%), and predictive power for various psychopathological domains. Lastly, we found a group of distinct syndromes were 4 - 7 (likelihood ratio) times more common in one construct than the other.

Comparison to prior analyses and tests of diagnostic accuracy

The majority of prior work examining psychopathology with multivariate statistics was conducted a number of years ago. The original discriminant function analysis by Kendell et al was undertaken on 292 patients from the general psychotic sample, with 38 clinical and historical predictors to construct the function, with 91% of cases correctly classified. They were unable to draw firm conclusions, based on lack of clear bimodality (Kendell and Gourlay, 1970). A re-analysis of a proportion of this sample, with the addition of functional outcome (over an average of 5.6 years) entered into the discriminant function produced a clear bimodal distribution, 96% of all patients correctly classified, 8 variables used to create the discriminant function (Brockington et al., 1979). Similar results have been found in analyses of IPSS studies (which also used PSE) (Carpenter et al.,

1973). The PSE was also used to assess discriminative properties of symptoms seen in patients seen by Kraepelin in 1908, diagnosed with either dementia praecox or manic-depressive insanity (Jablensky et al., 1993). Here, discriminant function and cluster analyses separated out both diagnostic constructs, which resembled ICD 9 and DSM IIR diagnostic categories. It should be noted that these techniques measured differences at group, and not at individual level.

Other findings have focused mainly on first rank symptoms (FRS)(Nordgaard et al., 2008).

The role of first rank symptoms

In our study, nuclear syndrome (first rank symptoms - FRS), was almost 3 times more common in schizophrenia compared to affective psychosis (see Table 1). This is relevant, given renewed interest in presence of FRS; ICD 11 proposes to take this out (Heinz et al., 2016; Lawrie et al., 2016). Our findings contrast with other studies, notably Peralta and Cuesta (Peralta and Cuesta, 1999), who found small differences in FRS between schizophrenia, using Feighner, DSM III-R narrow and broad criteria and other psychotic disorders, with likelihood ratios of 1 to 3.

Nordgaard et al examining prevalence of FRS identified methodological failings in prior studies. These included heterogeneity of populations, lack of clear definitions for schizophrenia (18 of 39 studies), insufficient sampling, interview/rating system, lack of comparison within FRS (e.g. lack of ego boundaries and auditory hallucinations), lack of reliable measures, and a mixture of illness variables (Nordgaard et al., 2008). The only measure of these in which our study falls short is demonstrable reliability, though the interviewers (DCGO, ECJ) were experienced clinicians, and had worked on similar projects together for

a number of years. A Cochrane review of FRS in schizophrenia (Soares-Weiser et al., 2015), utilizing data from 16 studies (4070 participants) found FRS differentiated schizophrenia from other types of psychosis with a sensitivity of 58.0% (95% CIs 50.3% to 65.3%) and specificity of 74.7% (95% CIs 65.2% to 82.3%). When DSM III operational criteria were used as part of the reference standard, in 4 studies, sensitivity was 64.8% (95% CIs 54.3-74) and specificity 64.2% (95% CIs 52.8-74.2). This review also commented on aspects of study quality, noting risk of bias regarding patient selection, use of index test and reference standard as well as blinding of those conducting the tests not being reported.

Strengths and limitations

Although completed a number of years ago, this dataset has a number of strengths, and it is unlikely that a similar dataset will be readily available to address this specific research question. The instrument used (PSE) is particularly thorough in eliciting symptoms of psychosis, and the degree of expertise of the raters is worth noting, as psychopathological domains may be adversely affected in those who do not have adequate expertise, with effects on data quality. The PSE and the DSM III diagnostic criteria were administered independent of each other, at different time points. The PSE assessment was blind to diagnostic category, and the DSM III diagnosis was made from case-notes approximately two and a half years later, thereby reducing observer bias. The time-consuming nature of the PSE, and the configuration of psychiatric services at the time of the original study (inpatient care for those with significant mental illness) is one of the strengths of the current analysis. Although not strictly a defined catchment-area population, this group of patients was fairly homogenous in terms of ethnicity, and referrals from the North

London area, and referred from both primary and secondary care, before the advent of community services, with a small degree of illicit substance use (E. C. Johnstone et al., 1992).

Approaches that search through a given dataset to find a model of the relationship tend to over-fit the data . The machine learning technique we have utilized overcomes this problem, because we have validated the model in an independent (from the training) test dataset. In addition, the application of this technique allows inference at the individual, rather than group level statistics used previously. Lastly, although our primary analysis was concerned with validity of the two constructs (diagnosed by DSM III), we were also able to show a discrimination, even when accounting for mood symptoms (using the data-driven hierarchical clustering approach), suggesting a more nuanced symptomatic difference in psychopathology.

Our study has a number of limitations. The diagnostic criteria applied to the sample was DSM III, which has important differences from DSM IV and DSM V (see below), Data on outcome was unavailable for the current re-analysis, though the question we asked pertained specifically to clinical syndromes, as opposed to being able to differentiate the two disorders by treatment response or outcome. As a diagnostic instrument the PSE, whilst covering the breadth of psychotic symptoms, does not fully cover mood symptoms. This limitation in the instrument did not, however, prevent us from being able to discriminate both constructs, despite being unable to tap into the full gamut of affective symptoms. Whilst one could conceivably state that the classification into DSM III diagnoses by those who had completed PSEs over two years previously could be a source of potential bias, the

modest agreement between PSE and DSM III diagnoses in the original paper (Cohen's kappa=0.49) argues against this.

DSM III diagnostic criteria compared to DSM V

The operationalizing of diagnostic criteria in DSM III represented a significant advance to prior criteria (Kendler, 2016), and makes the distinction between schizophrenia and affective psychosis on the basis of presence of mood symptoms. This broad discrimination remains in DSM V, differences existing in inclusion criteria for schizophrenia, such as change in volition. Another major difference is the narrower criteria for schizoaffective disorder, which probably limited the number of schizoaffective diagnoses made in this sample, giving us inadequate power to detect meaningful analyses for this construct.

Genetic findings and how they relate to nosological distinction between the affective psychoses and schizophrenia.

Whilst a comprehensive discussion of the genetic underpinnings of affective psychosis and schizophrenia is outwith the scope of this article, it is worth noting both the genetic overlap and distinctions between the two constructs.

A means of understanding the contribution of genetics to nosology has been shared genetic and environmental risk factors. In the case of schizophrenia this could be considered within a neurodevelopmental framework, with shared causality between genes and environment, including intellectual disability, autism, attention deficit hyperactivity disorder (ADHD), schizophrenia and possibly major affective syndromes (Owen, 2012). Notably, some studies indicate a link between schizophrenia and autism, though not bipolar disorder (Carroll and Owen, 2009).

The neurodevelopmental trajectories of schizophrenia and affective psychosis do appear to differ (Payá et al., 2013). Of the 108 associated loci picked up in the 2014 schizophrenia genome-wide association study (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), a number have not been identified in bipolar disorder.

The “point of rarity” debate

The algorithm we used could not classify individuals into schizophrenia or affective psychosis with 100 % accuracy, though the 83% accuracy we report is well above chance. The reduction of the true positive rate of affective psychosis on removal of affective syndromes from the model, suggests symptoms contributing to affective psychosis are perhaps distinct from those contributing to a diagnosis of schizophrenia. While accepting an overlap of symptoms between the two diagnoses, our data suggest distinguishable features that could help classify individuals into a distinct category. The results of the original discriminant function analysis prompted the notion that no “point of rarity” exists between affective psychosis and schizophrenia, and that nosological boundaries of the two disorders are indistinct (Kendell and Jablensky, 2003). While our aim was not to demonstrate a point of rarity between the two entities, our findings suggest the overlap between conditions alone may be insufficient to abandon the point of rarity concept. In other words, whether or not a statistical “point of rarity” exists, a statistical disparity does appear to exist between these constructs, and it is possible to differentiate the conditions with reasonable accuracy, at an individual level. We think this has clinical utility, in guiding management of affective and non-affective psychoses.

Conclusion

Acknowledging difficulties inherent within psychiatric classification, continued criticisms regarding the symptomatic distinction between the affective psychoses and schizophrenia seems premature. Bearing in mind recent initiatives proposing to combine classification with basic science research and transdiagnostic approaches (Insel et al., 2010; Jauhar et al., 2017), we would suggest that being able to discriminate the nature of psychotic disorder on the basis of certain syndromes is still of clinical importance, and that the presence of some psychopathological syndromes is as relevant today as they were over a century ago.

Competing interests

The authors declare no competing interests.

Acknowledgements

We would like to thank Professor Andrew McIntosh, Professor Gordon Murray and Professor Allan Young, who advised on the statistical analysis of an earlier version of the manuscript.

Dr Jauhar is funded by a JMAS Sim Fellowship from the Royal College of Physicians, Edinburgh and National Institute for Health Research Biomedical Research Centre at South London and Maudsley National Health Service Foundation Trust and King's College, London.

References

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders, DSM Library. American Psychiatric Association.

Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Studerus, E., Casalicchio, G., Jones, Z.M., 2016. mlr: Machine Learning in R. *Journal of Machine Learning Research* 17, 1-5.

Brockington, I., Kendell, R., Wainwright, S., Hillier, V., Walker, J., 1979. The distinction between the affective psychoses and schizophrenia. *The British Journal of Psychiatry* 135, 243–248. <https://doi.org/10.1192/bjp.135.3.243>

Carpenter, W.T., Strauss, J.S., Muleh, S., 1973. Are There Pathognomonic Symptoms in Schizophrenia?: An Empiric Investigation of Schneider's First-Rank Symptoms. *Arch Gen Psychiatry* 28, 847–852. <https://doi.org/10.1001/archpsyc.1973.01750360069010>

Carroll, L.S., Owen, M.J., 2009. Genetic overlap between autism, schizophrenia and bipolar disorder. *Genome Medicine* 1, 102. <https://doi.org/10.1186/gm102>

CRADDOCK, N., OWEN, M.J., 2007. Rethinking psychosis: the disadvantages of a dichotomous classification now outweigh the advantages. *World Psychiatry* 6, 84.

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* 33, 1–22.

Guloksuz, S., Os, J. van, 2018. The slow death of the concept of schizophrenia and the

painful birth of the psychosis spectrum. *Psychological Medicine* 48, 229–244.
<https://doi.org/10.1017/S0033291717001775>

Heinz, A., Voss, M., Lawrie, S.M., Mishara, A., Bauer, M., Gallinat, J., Juckel, G., Lang, U., Rapp, M., Falkai, P., Strik, W., Krystal, J., Abi-Dargham, A., Galderisi, S., 2016. Shall we really say goodbye to first rank symptoms? *Eur. Psychiatry* 37, 8–13.
<https://doi.org/10.1016/j.eurpsy.2016.04.010>

Insel, T.R., Cuthbert, B.N., Garvey, M.A., Heinssen, R.K., Pine, D.S., Quinn, K.J., Sanislow, C.A., Wang, P.S., 2010. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *American Journal of Psychiatry* 167, 748–751.

Jablensky, A., 2010. The diagnostic concept of schizophrenia: its history, evolution, and future prospects. *Dialogues in clinical neuroscience* 12, 271.

Jablensky, A., Hugler, H., Von Cranach, M., Kalinov, K., 1993. Kraepelin revisited: a reassessment and statistical analysis of dementia praecox and manic-depressive insanity in 1908. *Psychological Medicine* 23, 843–858.

Jauhar, S., Nour, M.M., Veronese, M., Rogdaki, M., Bonoldi, I., Azis, M., Turkheimer, F., McGuire, P., Young, A.H., Howes, O.D., 2017. A Test of the Transdiagnostic Dopamine Hypothesis of Psychosis Using Positron Emission Tomographic Imaging in Bipolar Affective Disorder and Schizophrenia. *JAMA Psychiatry*.
<https://doi.org/10.1001/jamapsychiatry.2017.2943>

Johnstone, E.C., Frith, C., Crow, T., Owens, D., Done, D., Baldwin, E., Charlette, A., 1992. The Northwick Park “functional” psychosis study: diagnosis and outcome. *Psychological Medicine* 22, 331–346.

Johnstone, E.C., Frith, C.D., Crow, T.J., Owens, D.G.C., Done, D.J., Baldwin, E.J., Charlette, A., 1992. The Northwick Park “Functional” Psychosis Study: Diagnosis and Outcome. *Psychological Medicine* 22, 331–346. <https://doi.org/10.1017/S0033291700030270>

Kendell, R., Jablensky, A., 2003. Distinguishing between the validity and utility of psychiatric diagnoses. *Am J Psychiatry* 160, 4–12.

Kendell, R.E., Brockington, I.F., Leff, J.P., 1979. Prognostic implications of six alternative definitions of schizophrenia. *Arch. Gen. Psychiatry* 36, 25–31.

Kendell, R.E., Gourlay, J., 1970. The Clinical Distinction between the Affective Psychoses and Schizophrenia. *BJP* 117, 261–266. <https://doi.org/10.1192/bjp.117.538.261>

Kendler, K.S., 2016. Phenomenology of Schizophrenia and the Representativeness of Modern Diagnostic Criteria. *JAMA Psychiatry* 73, 1082–1092.
<https://doi.org/10.1001/jamapsychiatry.2016.1976>

Kraepelin, E., 1987. *Dementia praecox*. Cambridge University Press.

Lawrie, S.M., Hall, J., McIntosh, A.M., Owens, D.G.C., Johnstone, E.C., 2010. The “continuum of psychosis”: scientifically unproven and clinically impractical. *Br J Psychiatry* 197, 423–425. <https://doi.org/10.1192/bjp.bp.109.072827>

Lawrie, S.M., O’Donovan, M.C., Saks, E., Burns, T., Lieberman, J.A., 2016. Improving classification of psychoses. *The Lancet Psychiatry* 3, 367–374.

[https://doi.org/10.1016/S2215-0366\(15\)00577-5](https://doi.org/10.1016/S2215-0366(15)00577-5)

Linscott, R.J., Allardyce, J., Van Os, J., 2010a. Seeking verisimilitude in a class: a systematic review of evidence that the criterial clinical symptoms of schizophrenia are taxonic. *Schizophrenia bulletin* 36, 811.

Linscott, R.J., Lenzenweger, M.F., Os, J., 2010b. Continua or classes? Vexed questions on the latent structure of schizophrenia. *Advances in Schizophrenia Research* 2009 333–355.

Manton, K.G., Korten, A., Woodbury, M.A., Anker, M., Jablensky, A., 1994. Symptom profiles of psychiatric disorders based on graded disease classes: an illustration using data from the WHO International Pilot Study of Schizophrenia. *Psychological medicine* 24, 133–144.

Mechelli, A., Lin, A., Wood, S., McGorry, P., Amminger, P., Tognin, S., McGuire, P., Young, J., Nelson, B., Yung, A., 2016. Using clinical information to make individualized prognostic predictions in people at ultra high risk for psychosis. *Schizophr. Res.*
<https://doi.org/10.1016/j.schres.2016.11.047>

Mercaldo, N.D., Lau, K.F., Zhou, X.H., 2007. Confidence intervals for predictive values with an emphasis to case-control studies. *Stat Med* 26, 2170–2183.
<https://doi.org/10.1002/sim.2677>

Möller, H., Jäger, M., Riedel, M., Obermeier, M., Strauss, A., Bottlender, R., 2011. The Munich 15-year follow-up study (MUFUSSAD) on first-hospitalized patients with schizophrenic or affective disorders: Assessing courses, types and time stability of diagnostic classification. *European Psychiatry* 26, 231–243.

Murray, V., McKEE, I., Miller, P.M., Young, D., Muir, W.J., Pelosi, A.J., Blackwood, D.H.R., 2005. Dimensions and classes of psychosis in a population cohort: a four-class, four-dimension model of schizophrenia and affective psychoses. *Psychological Medicine* 35, 499–510. <https://doi.org/10.1017/S0033291704003745>

Nordgaard, J., Arnfred, S.M., Handest, P., Parnas, J., 2008. The Diagnostic Status of First-Rank Symptoms. *Schizophr Bull* 34, 137–154. <https://doi.org/10.1093/schbul/sbm044>

Organization, W.H., 1993. The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research. World Health Organization.

Owen, M.J., 2012. Implications of Genetic Findings for Understanding Schizophrenia. *Schizophr Bull* 38, 904–907. <https://doi.org/10.1093/schbul/sbs103>

Payá, B., Rodríguez-Sánchez, J.M., Otero, S., Muñoz, P., Castro-Fornieles, J., Parellada, M., Gonzalez-Pinto, A., Soutullo, C., Baeza, I., Rapado-Castro, M., Sáenz-Herrero, M., Moreno, D., Arango, C., 2013. Premorbid impairments in early-onset psychosis: Differences between patients with schizophrenia and bipolar disorder. *Schizophrenia Research* 146, 103–110.
<https://doi.org/10.1016/j.schres.2013.01.029>

Peralta, V., Cuesta, M.J., 1999. Diagnostic Significance of Schneider's First-Rank Symptoms in Schizophrenia. Comparative Study Between Schizophrenic and Non-Schizophrenic Psychotic Disorders. *BJP* 174, 243–248. <https://doi.org/10.1192/bjp.174.3.243>

Pomarol-Clotet, E., Salvador, R., Murray, G., Tandon, S., McKenna, P.J., 2010. Are There Valid Subtypes of Schizophrenia? A Grade of Membership Analysis. *Psychopathology* 43,

53–62. <https://doi.org/10.1159/000260044>

Samuel, A.L., 1959. Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development* 3, 210–229. <https://doi.org/10.1147/rd.33.0210>

Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511, 421–427. <https://doi.org/10.1038/nature13595>

Soares-Weiser, K., Maayan, N., Bergman, H., Davenport, C., Kirkham, A.J., Grabowski, S., Adams, C.E., 2015. First rank symptoms for schizophrenia, in: *The Cochrane Collaboration* (Ed.), *Cochrane Database of Systematic Reviews*. John Wiley & Sons, Ltd, Chichester, UK.

Suzuki, R., Shimodaira, H., 2006. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22, 1540–1542. <https://doi.org/10.1093/bioinformatics/btl117>

Tandon, R., Keshavan, M.S., Nasrallah, H.A., 2008. Schizophrenia, “Just the Facts”: what we know in 2008 part 1: overview. *Schizophr. Res.* 100, 4–19. <https://doi.org/10.1016/j.schres.2008.01.022>

Wing, J.K., Cooper, J.E., Sartorius, N., 1974. *Measurement and classification of psychiatric symptoms; An instruction manual for the PSE and Catego program*. Cambridge U Press.

Figures and Tables

Figure 1 Study Flowchart

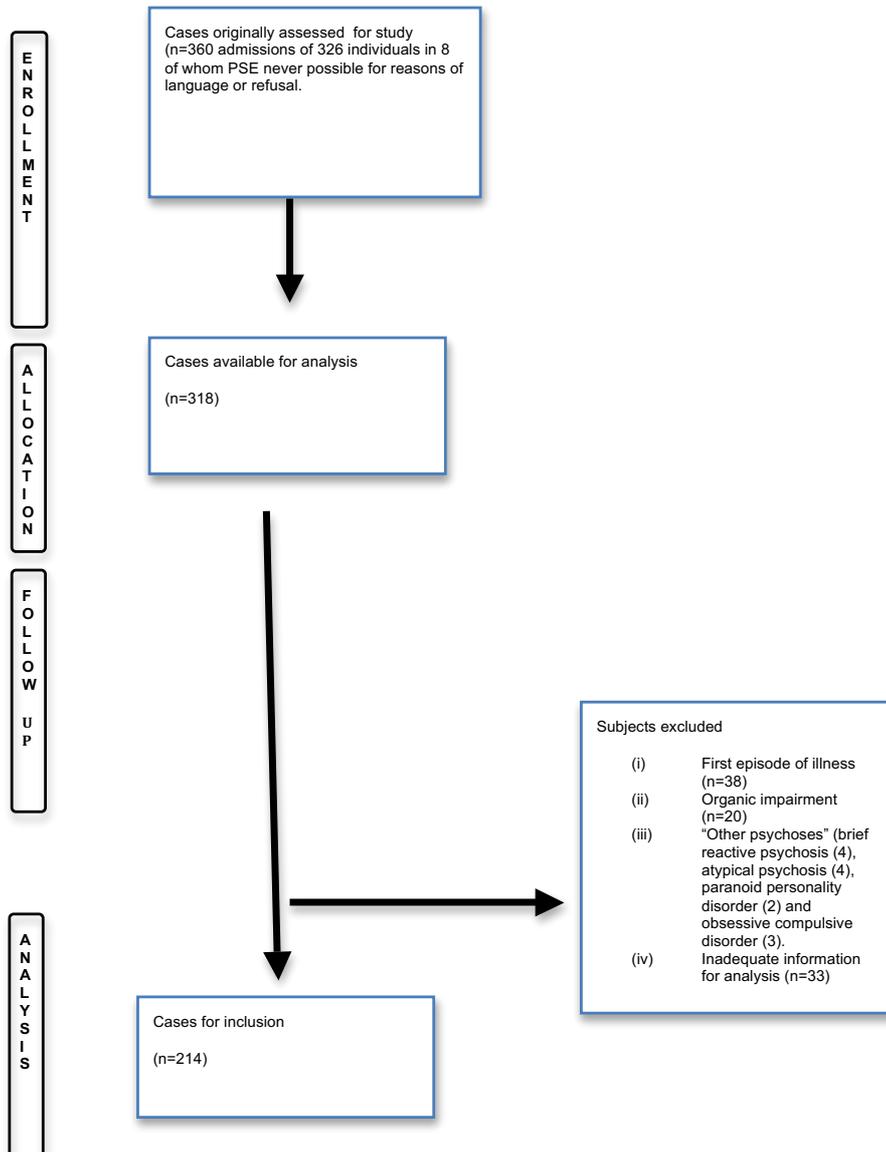


Figure 2. Hierarchical clustering analysis of PSE syndromes

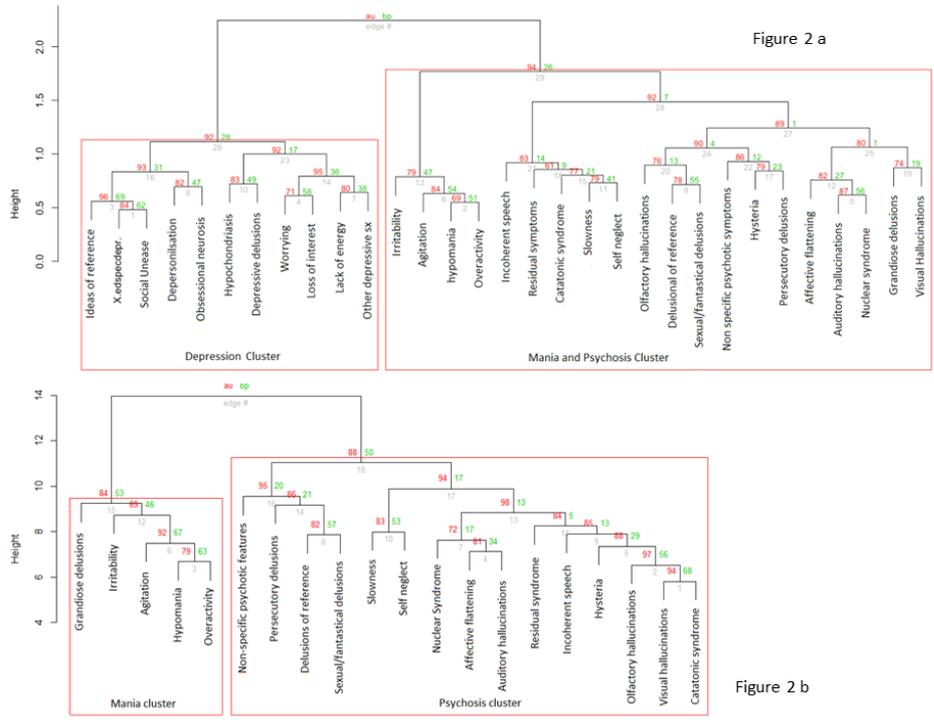


Figure 3: ROC curve showing the area under the curve of PSE in distinguishing non-affective from affective psychosis in the test dataset.

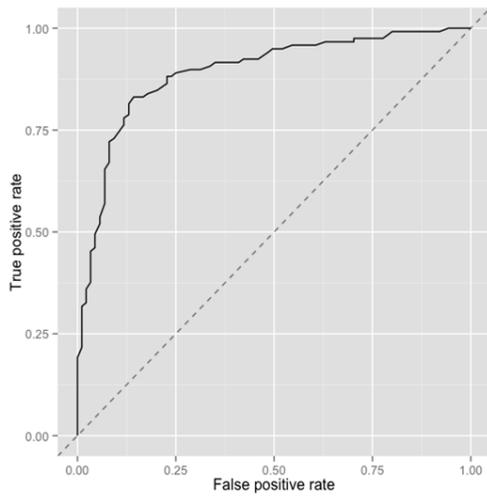


Figure a

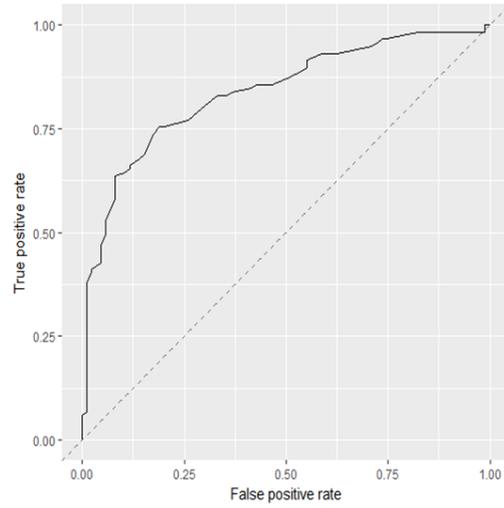


Figure b

Table 1 Logistic regression analysis of relationship between non-affective syndromes at initial presentation and diagnosis of schizophrenia (compared to affective psychosis).

<i>Variable Name</i>	<i>Odds Ratio (OR) of</i>		
	<i>schizophrenia</i>	<i>95% CI of OR</i>	<i>p Value</i>
Hysteria	0.55	0.15 - 1.96	0.355
Affective flattening	12.24	2.16 - 69.38	0.005
Auditory			
hallucinations	29.82	4.90 - 181.50	<0.001
Persecutory delusions	3.86	1.53 - 9.73	0.004
Delusions of reference	3.45	1.13 - 10.50	0.029
Sexual / fantastical			
delusions	1.03	0.38 - 2.82	0.953
Visual hallucinations	0.85	0.19 - 3.89	0.835
Olfactory hallucinations	1.77	0.46 - 6.84	0.409
Slowness	0.72	0.24 - 2.19	0.566
Non-specific feat.			
psychosis	1.59	0.67 - 3.77	0.290
Depersonalisation	0.50	0.13 - 2.00	0.329
Agitation	0.37	0.14 - 0.99	0.047
Self-neglect	1.76	0.63 - 4.90	0.283

Ideas of reference	0.90	0.25 - 3.18	0.864
Lack of energy	0.47	0.16 - 1.44	0.186
Worrying	1.13	0.35 - 3.60	0.837
Social unease	3.39	1.05 - 10.95	0.042
Loss of interest	0.44	0.16 - 1.23	0.117
Hypochondriasis	1.68	0.57 - 4.95	0.345
Nuclear syndrome	3.26	1.25 - 8.47	0.015
Catatonic syndrome	3.59	0.59 - 21.81	0.165
Incoherent Speech	3.13	1.07 - 9.18	0.038
Residual syndrome	0.80	0.29 - 2.25	0.677
Obsessional neurosis	0.42	0.10 - 1.82	0.245

202 observations, 177 error degrees of freedom. Chi-squared statistic vs. constant model: 122, $p < 0.001$. Adjusted R-squared=0.448. Significant variables highlighted in bold.

	Estimate	Standard Error	Odds Ratio	p	Lower bound	Upper bound
Hysteria.	-0.47	0.56	0.62	0.40	0.20	1.88
Affective flattening.	2.54	0.82	12.69	0.002	2.49	64.52
Auditory hallucinations.	2.81	0.82	16.65	< .001	3.32	83.45

	Estimate	Standard Error	Odds Ratio	p	Lower bound	Upper bound
Persecutory delusions.	1.20	0.41	3.35	0.003	1.49	7.51
Delusions of reference.	0.86	0.51	2.38	0.09	0.86	6.58
Sexual and fantastical delusions.	0.24	0.45	1.27	0.59	0.52	3.10
Visual hallucinations.	-0.33	0.67	0.71	0.61	0.19	2.68
Olfactory hallucinations.	0.47	0.61	1.60	0.44	0.48	5.30
Slowness.	-0.29	0.48	0.74	0.54	0.28	1.93
Non-specific features of psychosis.	0.59	0.40	1.82	0.13	0.82	4.02
Self-neglect.	0.62	0.46	1.86	0.18	0.74	4.68
Nuclear syndrome.	0.97	0.44	2.65	0.02	1.10	6.38
Catatonic syndrome.	1.30	0.88	3.67	0.14	0.64	20.97
Incoherent Speech.	0.95	0.51	2.59	0.06	0.94	7.09
Residual syndrome.	-0.21	0.49	0.80	0.65	0.30	2.11

Table 2 Logistic regression analysis of relationship between non-affective syndromes at initial presentation and diagnosis of schizophrenia (compared to affective psychosis). 202 observations, 186 error degrees of freedom. Chi-squared

statistic vs. constant model: 106.47, $p < 0.001$. Adjusted R-squared 0.55. Significant variables highlighted in bold.