Predicting one-year outcome in first episode psychosis using machine learning

Citation for published version:

Digital Object Identifier (DOI):
10.1371/journal.pone.0212846

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
PLoS ONE

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Predicting one-year outcome in first episode psychosis using machine learning

Samuel P. Leighton, Rajeev Krishnadas, Kelly Chung, Alison Blair, Susie Brown, Suzy Clark, Kathryn Sowerbutts, Matthias Schwannauer, Jonathan Cavanagh, Andrew I. Gumley

1 Institute of Health and Wellbeing, University of Glasgow, Glasgow, United Kingdom, 2 Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, United Kingdom, 3 ESTEEM First Episode Psychosis Service, NHS Greater Glasgow and Clyde, Glasgow, United Kingdom, 4 Department of Clinical & Health Psychology, University of Edinburgh, Edinburgh, United Kingdom

‡ These authors are joint senior authors on this work.
* Rajeev.Krishnadas@glasgow.ac.uk

Abstract

Background

Early illness course correlates with long-term outcome in psychosis. Accurate prediction could allow more focused intervention. Earlier intervention corresponds to significantly better symptomatic and functional outcomes. Our study objective is to use routinely collected baseline demographic and clinical characteristics to predict employment, education or training (EET) status, and symptom remission in patients with first episode psychosis (FEP) at one-year.

Methods and findings

83 FEP patients were recruited from National Health Service (NHS) Glasgow between 2011 and 2014 to a 24-month prospective cohort study with regular assessment of demographic and psychometric measures. An external independent cohort of 79 FEP patients were recruited from NHS Glasgow and Edinburgh during a 12-month study between 2006 and 2009. Elastic net regularised logistic regression models were built to predict binary EET status, period and point remission outcomes at one-year on 83 Glasgow patients (training dataset). Models were externally validated on an independent dataset of 79 patients from Glasgow and Edinburgh (validation dataset). Only baseline predictors shared across both cohorts were made available for model training and validation. After excluding participants with missing outcomes, models were built on the training dataset for EET status, period and point remission outcomes and externally validated on the validation dataset. Models predicted EET status, period and point remission with receiver operating curve (ROC) area under the curve (AUC) performances of 0.876 (95%CI: 0.864, 0.887), 0.630 (95%CI: 0.612, 0.647) and 0.652 (95%CI: 0.635, 0.670) respectively. Positive predictors of EET included baseline EET and living with spouse/children. Negative predictors included higher PANSS suspiciousness, hostility and delusions scores. Positive predictors for symptom remission included living with spouse/children, and affective symptoms on the Positive and Negative Syndrome Scale (PANSS). Negative predictors of remission included passive social...
withdrawal symptoms on PANSS. A key limitation of this study is the small sample size (n) relative to the number of predictors (p), whereby p approaches n. The use of elastic net regularised regression rather than ordinary least squares regression helped circumvent this difficulty. Further, we did not have information for biological and additional social variables, such as nicotine dependence, which observational studies have linked to outcomes in psychosis.

Conclusions and relevance
Using advanced statistical machine learning techniques, we provide the first externally validated evidence, in a temporally and geographically independent cohort, for the ability to predict one-year EET status and symptom remission in individual FEP patients.

Introduction
Initial clinical presentation and early illness course correlates with long term outcome in first episode psychosis (FEP).[1] The ability to accurately predict outcome at an individual level would allow more focussed intervention. For patients, a meaningful outcome is often more than simple symptom remission but optimising developmental pathways in early psychosis including vocational and educational outcomes.[2–4]

FEP most often presents at the critical stage in a young person’s life when community, societal roles, educational and vocational achievement are being forged. Consequently, its onset triggers a precipitous decline in education and employment. Missing out on such vocational opportunities, as enshrined in Article 23 of the Universal Declaration of Human Rights[5], impairs not only financial independence but also societal inclusion, the forming of relationships and self-actualisation.[6]

Those with FEP want to work. Evidence suggests that finding employment is more important than any specific mental health intervention.[7] Yet less than a quarter of those with severe mental illness such as schizophrenia receive vocational rehabilitation.[8,9] Clinicians’ attitudes towards their patients’ ability to return to work and their associated estimation of risks are often ambivalent; this is reinforced by a continued decline in employment rates in the years following contact with mental health services.[10]

Attitudes are changing with a new focus on employment, education and training (EET) recognised by the Meaningful Lives international consensus statement for FEP.[11] Successful intervention strategies exist, such as the Individual Placement and Support approach in FEP, which have been evidenced to show employment and education rates of 69% as compared with 35% for controls.[10] If we can correctly identify those with poor EET outcomes at their initial presentation, we could apply such vocational interventions at an earlier stage. Existing evidence suggests this would have a much greater chance of success.[12,13]

At present, our knowledge of factors, which predict outcome in FEP is incomplete. A recent meta-analysis by Lally et al provides the first robust evidence of remission and recovery outcomes in FEP but was unable to establish the key clinical or demographic factors, which discriminated between patients. Specifically, it did not replicate an earlier meta-analysis showing an association between longer duration of untreated psychosis (DUP) and worse outcome in FEP.[14,15] A systematic review identified poor premorbid adjustment, history of developmental disorder, greater symptom severity at baseline and longer DUP to be the most replicated predictors of poor clinical, functional, cognitive, and biological outcomes in early onset
psychosis. Within the Scottish population specifically, socioeconomic deprivation and ethnicity have been shown to be risk factors for developing psychosis, while substance misuse, longer first admission, younger age of onset and male gender increased the risk of poor long-term outcome. However, such group level differences cannot be extrapolated to individuals—the 'ecological fallacy'—nor can observational results be readily equated with causation or accurate predictions.

Advanced machine learning techniques have potential to revolutionise medicine by looking at causation and the prediction of individual patient outcome. Within psychiatry, machine learning has been already applied to symptomatic and neuroimaging data to classify individuals into diagnostic categories and to predict response to ECT, with high levels of accuracy. Koutsouleris et al employed machine learning to predict 4 and 52-week outcome (Global Assessment of Function \( \geq 65 \)) in FEP to a 75% and 73.8% test-fold balanced accuracy on repeated nested internal cross-validation. The authors suggest that before employing a machine learning model "into real-world care, replication is needed in external first episode samples".

Because of the practicalities surrounding the requirement to recruit an additional and independent cohort of patients, few studies assess predictive model generalisability via temporal and/or geographical validation. External validation in a plausibly related population is a considerably stronger test of predictive models. It assesses model transportability to an untested situation rather than simply reproducibility alone. As outlined by Koutsouleris et al, this is considered essential before applying the predictive model to clinical practice.

Our study is based on the hypothesis that it is possible to predict outcome in terms of employment education or training (EET) status or symptom remission at one-year in FEP using baseline demographic and clinical psychometric biomarkers. We are not aware of any study to date, which has employed predictive modelling of EET status in a FEP cohort. We then assess the generalisability of our predictive model on an external cohort of patients.

Additionally, we seek to find the relative importance of the individual variables in contributing to prediction performance. These could have clinical relevance and may inform future research. Focussing on patients with FEP, who have not had extensive psychotherapeutic or pharmacological interventions, helps mitigate against potential confounders.

**Methods**

The study was approved by the West of Scotland Research Ethics Committee. Reference number: 11/AL/0247 and R and D approval (GN11CP130). Written informed consent was obtained from all participants in the study.

**Participants and study design**

The Compassionate Recovery: Individualised Support in early Psychosis (CR:ISP) study was a 24-month non-randomised prospective study of individuals with FEP. An Integrated Care Pathway for FEP was implemented which facilitated regular routine assessment of demographic and psychometric measures at time 0, month 3, month 6 and month 12.

Recruitment took place in mental health services in the NHS Greater Glasgow & Clyde (NHSGGC) health board between 2011 and 2014. To be considered for inclusion participants had to be: (a) in-patients or out-patients with (b) first presentation to mental health services for psychosis, (c) ICD-10 diagnosis of non-organic psychosis. 83 participants were entered into the study. The CR:ISP patients formed the training dataset.
The validation dataset was formed of 79 FEP patients recruited to an earlier study (1 September 2006 to 31 August 2009), which took place in mental health services in the NHS in Glasgow and in Edinburgh. Demographic and psychometric measures were assessed at time 0, month 6 and month 12. Participants and study design have been described previously (see S1 File).[28]

**Predictor and outcome measures**

Baseline predictors shared across both cohorts were made available for model training and external validation. Demographic predictors included admission to hospital, age, citizenship, educational attainment, ethnicity, gender, household composition, EET at baseline, parental status, relationship status, accommodation, alcohol use, and recreational drug use status. Psychometric clinical predictors included individual PANSS items[29] and ordinalised depression rating (training cohort used Hospital Anxiety and Depression Scale (HADS), validation cohort used Beck’s Depression Inventory II (BDI-II)—scores were categorised as none, mild, moderate or severe according to published cut offs.[30,31]

The one-year binary outcome measures included EET status at one-year, PANSS point remission (meeting Andreasen PANSS criteria at month 12), and period remission (meeting Andreasen PANSS criteria at both month 6 and month 12). Andreasen et al defined remission as scores of less than or equal to 3 in PANSS items P1 Delusions, P2 Conceptual Disorganisation, P3 Hallucinatory Behaviour, N1 Blunted Affect, N4 Apathetic Social Withdrawal, N6 Lack of Spontaneity and G9 Unusual Thought Content, present for a period of at least 6 months.[32]

**Statistical analysis**

All statistical analyses were carried out within the R programming environment.[33]

Between-group (training versus validation) differences were tested using Welch’s independent t-test for continuous variables and Pearson’s chi-squared or Fisher’s exact test for categorical variables. Bonferroni correction was employed for multiple comparisons.

Machine learning analysis was carried out using the ‘Caret’ package.[34] R code is available in S2 File. The anonymised training and validation datasets are available in S3 File.

During pre-processing, data were centred and scaled, variables with zero variance and near-zero variance removed and missing data imputed using k (5) nearest neighbour imputation, prior to model construction. In psychiatric research, data are commonly seen as ‘missing not at random’. For example, with drop-out more frequent in those who have relapsed. Ignoring this leads to systematic attrition bias in any inference drawn, in addition to a loss of power. Imputation increases power and reduces bias.[35,36]

A logistic regression model was fit by elastic net regularisation with variable selection in ‘Caret’ with the ‘glmnet’ package.[37] This was following initial scoping utilising alternative classifiers including linear and radial support vector machines, random forest, naïve Bayes and linear discriminant analysis. Glimnet fits a generalized linear model via penalised maximum likelihood.[38] The objective function for the penalised logistic regression uses the negative binomial log-likelihood:

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

The alpha ($\alpha$) hyperparameter determines the balance between ridge ($\alpha = 0$) and lasso ($\alpha = 1$) penalty and the lambda ($\lambda$) hyperparameter determines the amount of penalty. Elastic net regression improves upon ordinary least squares via regularisation (shrinkage) of the estimated beta coefficients ($\beta$). This results in superior performance when number of predictors...
(p) approaches the number of subjects (N) (in such circumstances there is no unique least squares estimate), or in the presence multicollinearity by offsetting a small amount of bias with large reductions in variance. Further, a consequence of regularisation is feature selection, which improves model interpretability.[39]

We used n (5) fold repeated (100 times) cross validation to train and tune our model over a grid of alpha and lambda hyperparameters on our training dataset (see S1 File). All splits were balanced by outcome class. The model was refit on the whole training set (with the best performing hyperparameters) to calculate its standardised beta coefficients. The exponentials of which, the odds ratios, are presented.

We estimated the discriminative performance of the models using receiver operating curve (ROC) area under the curve (AUC). ROCAUC is independent of class distribution and selects for discriminative models with false and true positive rates significantly better than random chance.

External validity was established by using the model built on the training dataset to predict the probability of the outcome class and comparing it to the actual outcome class in the validation dataset, then, calculating the ROC AUC performance metric using the ‘pROC’ package. [40,41] The 95%CI of the ROCAUC was established based on U-statistic theory using the ‘clinfun’ package. [41,42] In addition, permutation testing was used to confirm significance, whereby, the actual ROCAUC was compared to its null ROCAUC distribution derived by testing the model on randomly permuted class outcomes, repeated 10000 times. The p value is the proportion of permuted values greater than or equal to the actual value.[43] The model accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) are presented based on the point closest to the top-left of the ROC plot. The formulae for these quantities have been described previously.[44]

The above procedure was repeated for each of the three dependent outcome variables.

Results

Table 1 summarises the predictor and outcome data for the training and validation cohorts. The only statistically significantly different predictor variable between the cohorts was accommodation. More people lived in rented accommodation in the training cohort, and more people lived with their family in the validation cohort (p < 0.001). There were no statistically significant differences in each of the three outcomes between the cohorts.

Table 2 summarises the EET status, point remission, and period remission models’ externally validated performance. Fig 1 show each model’s respective ROC curve–EET status (A), point remission (B) and period remission (C). All three models’ ROCAUC and 95%CI were significantly better than chance. The model predicting EET status had particularly high performance with a ROCAUC of 0.876 (95%CI: 0.864, 0.887; p = <0.001).

Figs 2–4 summarise the feature selected predictors and their odds ratios for the EET, point remission and period remission models.

Discussion

We believe our study to be the first externally validated evidence, in a temporally and geographically independent cohort, for predictive modelling in FEP at an individual patient level. This builds on existing studies of group-level differences and the elegant work of Koutsouleris et al, which outlined the first internally validated evidence of the ability to predict functional outcome in individual patients with FEP.[25]

Our results demonstrate the ability to predict both symptom remission and more importantly, functioning (in employment, education or training). The performance of our EET
Table 1. Characteristics of training cohort and validation cohort.

<table>
<thead>
<tr>
<th>Baseline Predictor</th>
<th>Training Cohort (n = 83)</th>
<th>Missing No. (%)</th>
<th>Validation Cohort (n = 79)</th>
<th>Missing No. (%)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accommodation—No. (%)</strong></td>
<td>Rented– 48 (59)</td>
<td>1 (1)</td>
<td>Rented– 20 (28)</td>
<td>8 (10)</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td></td>
<td>Private with family– 30 (37)</td>
<td></td>
<td>Private with family– 43 (61)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Private owner– 4(5)</td>
<td></td>
<td>Private owner– 4(6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Homeless– 0 (0)</td>
<td></td>
<td>Homeless– 2 (3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NFA– 0 (0)</td>
<td></td>
<td>NFA– 1 (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prison– 0 (0)</td>
<td></td>
<td>Prison– 1 (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age in Years—Mean (SD)</strong></td>
<td>25.22 (5.54)</td>
<td>0 (0)</td>
<td>24.64 (7.08)</td>
<td>3 (4)</td>
<td>0.6</td>
</tr>
<tr>
<td><strong>Alcohol Use—No. (%)</strong></td>
<td>58 (74)</td>
<td>5 (6)</td>
<td>58 (85)</td>
<td>11 (14)</td>
<td>0.1</td>
</tr>
<tr>
<td><strong>Citizenship—No. (%)</strong></td>
<td>UK– 72 (87)</td>
<td>0 (0)</td>
<td>UK– 74 (95)</td>
<td>1 (1)</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>Other– 5 (6)</td>
<td></td>
<td>Other– 3 (4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Asylum seeker– 4(5)</td>
<td></td>
<td>Asylum seeker– 0(0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Refugee– 2 (2)</td>
<td></td>
<td>Refugee– 1 (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Depression Rating—No. (%)</strong></td>
<td>None– 37 (50)</td>
<td>9 (11)</td>
<td>None– 16 (32)</td>
<td>29 (37)</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>Mild– 11 (15)</td>
<td></td>
<td>Mild– 10 (20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moderate– 13 (18)</td>
<td></td>
<td>Moderate– 11 (22)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Severe– 13 (18)</td>
<td></td>
<td>Severe– 13 (26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Drug Use—No. (%)</strong></td>
<td>42 (58)</td>
<td>11 (13)</td>
<td>37 (64)</td>
<td>21 (27)</td>
<td>0.5</td>
</tr>
<tr>
<td><strong>Highest Educational Attainment—No. (%)</strong></td>
<td>Before 16–14 (18)</td>
<td>3 (4)</td>
<td>Before 16–16 (23)</td>
<td>8 (10)</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>At 16–17 (21)</td>
<td></td>
<td>At 16–17 (24)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>17 to 18–21 (26)</td>
<td></td>
<td>17 to 18–20 (28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>College– 19 (23)</td>
<td></td>
<td>College– 11 (15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>University– 9 (11)</td>
<td></td>
<td>University– 7 (10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ethnicity—No. (%)</strong></td>
<td>White– 67 (81)</td>
<td>0 (0)</td>
<td>White– 74 (96)</td>
<td>2 (3)</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>Other– 16 (19)</td>
<td></td>
<td>Other– 3 (4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Gender—No. (%)</strong></td>
<td>Male– 55 (66)</td>
<td>0 (0)</td>
<td>Male– 54 (68)</td>
<td>0 (0)</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>Female– 28 (34)</td>
<td></td>
<td>Female– 25 (32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Hospital Admission at Baseline—No. (%)</strong></td>
<td>32 (39)</td>
<td>1 (1)</td>
<td>40 (51)</td>
<td>1 (1)</td>
<td>0.1</td>
</tr>
<tr>
<td><strong>Others in Household—No. (%)</strong></td>
<td>Parents &amp; siblings– 42 (51)</td>
<td>1 (1)</td>
<td>Parents &amp; siblings– 38 (55)</td>
<td>10 (13)</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>Alone– 15 (18)</td>
<td></td>
<td>Alone– 22 (32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Friends– 9 (11)</td>
<td></td>
<td>Friends– 4 (6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Siblings– 3 (4)</td>
<td></td>
<td>Siblings– 0 (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PANSS Total—Mean (SD)</strong></td>
<td>83.78</td>
<td>1 (1)</td>
<td>74.43</td>
<td>3 (4)</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>-24.62</td>
<td></td>
<td>-21.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Parent—No. (%)</strong></td>
<td>13 (16)</td>
<td>0 (0)</td>
<td>9 (12)</td>
<td>1 (1)</td>
<td>0.4</td>
</tr>
<tr>
<td><strong>Employment, Education or Training—No. (%)</strong></td>
<td>45 (56)</td>
<td>2 (2)</td>
<td>40 (52)</td>
<td>2 (3)</td>
<td>0.6</td>
</tr>
<tr>
<td><strong>Relationship Status—No. (%)</strong></td>
<td>Single– 65 (78)</td>
<td>0 (0)</td>
<td>Single– 69 (88)</td>
<td>1 (1)</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>Relationship– 12 (14)</td>
<td></td>
<td>Relationship– 5 (6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Married– 3 (4)</td>
<td></td>
<td>Married– 1 (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Separated– 3 (4)</td>
<td></td>
<td>Separated– 2 (2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Divorced– 0 (0)</td>
<td></td>
<td>Divorced– 1 (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>One-year Outcome</strong></td>
<td>Training Cohort No. (%)</td>
<td></td>
<td>Validation Cohort No. (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Period Remission</strong></td>
<td>33 / 67 (49)</td>
<td></td>
<td>33 / 64 (52)</td>
<td></td>
<td>0.8</td>
</tr>
<tr>
<td><strong>Point Remission</strong></td>
<td>40 / 71 (56)</td>
<td></td>
<td>46 / 67 (69)</td>
<td></td>
<td>0.1</td>
</tr>
<tr>
<td><strong>Employment, Education or Training Status</strong></td>
<td>32 / 75 (43)</td>
<td></td>
<td>27 / 67 (40)</td>
<td></td>
<td>0.8</td>
</tr>
</tbody>
</table>

* indicates Bonferroni corrected significance (p = 2.78×10^{-3}).
SD = standard deviation; No. = number of subjects.

https://doi.org/10.1371/journal.pone.0212846.t001
model was particularly robust, with an ability to accurately predict the one-year EET outcome in more than 85% of patients. This is clinically relevant, especially in the context of the 2010 international first episode psychosis “Meaningful Lives” consensus statement on unemployment as a source of significant social disability.[11]

Predictive statistics focus on the predictive performance of models on new ‘unseen’ data. This helps avoid the inherent problems with conventional statistical techniques, which only explore relationships within the study dataset. Few such relationships are externally validated and, of those that are, many lose significance.[45,46]

Because of the inherent feature selection of regularised logistic regression classifiers, which avoids overfitting in ill-conditioned regression problems where the number of variables is close to the sample size, our models are sparse and uniquely interpretable. Further, without a priori selection of variables but instead allowing the classifier to select features, we avoid the introduction of additional observer bias.[37,39]

The period remission model shares all the same predictors that are selected for the point remission model. However, the period remission model selects for additional predictors but with a reduced performance. This suggests possible overfitting on the training dataset.

Table 2. One-year EET status, point remission, and period remission models’ externally validated performance metrics.

<table>
<thead>
<tr>
<th>Performance Metric</th>
<th>Employment, Education or Training Status</th>
<th>Point Remission</th>
<th>Period Remission</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROC AUC</td>
<td>0.876 (95%CI: 0.864, 0.887) p = &lt;0.001</td>
<td>0.652 (95%CI: 0.635, 0.670) p = 0.048</td>
<td>0.630 (95%CI: 0.612, 0.647) p = 0.076</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.851</td>
<td>0.612</td>
<td>0.625</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.815</td>
<td>0.578</td>
<td>0.606</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.875</td>
<td>0.667</td>
<td>0.645</td>
</tr>
<tr>
<td>PPV</td>
<td>0.815</td>
<td>0.794</td>
<td>0.645</td>
</tr>
<tr>
<td>NPV</td>
<td>0.875</td>
<td>0.424</td>
<td>0.606</td>
</tr>
</tbody>
</table>

The 95%CI of the ROC AUC was established based on U-statistic theory, and significance level by permutation testing (n = 10000). Representative accuracy, sensitivity, specificity, positive predictive, and negative predictive values are presented in italics based on the point on the ROC curve closest to the top left. ROC AUC = Receiver Operating Characteristic Area Under the Curve; PPV = Positive Predictive Value; NPV = Negative Predictive Value; 95%CI = 95% Confidence Interval.

Fig 1. ROC curves for model of one-year EET status (A), point remission (B), and period remission (C). ROC = Receiver Operating Characteristic; EET = Employment, Education or Training.
whereby the signal to noise variable ratio captured by the model is reduced, hence poorer generalisability to the independent validation dataset.[47]

Focussing on the shared predictors for both point and period remission, higher scores on the individual PANSS items P4 Excitement and G6 Depression correspond to a positive outcome. Over a century later, this continues to lend credence to a central tenant of the Kraepelian dichotomy—that psychosis in the context of phenomenologically prominent affective presentations tends to have a better chance of recovery.[48] Recent research in this area is conflicting and hampered by the few prospective studies looking at the influence of affective symptoms on outcome.[49]

PANSS item N4 passive or apathetic social withdrawal predicts poor one-year remission outcomes, a so called negative symptom frequently found in the classical schizophrenia syndrome.[50] We have previously identified social and interpersonal factors as influencing outcome in FEP.[28,51] Such initial negative symptoms persist longer and are more difficult to

Fig 2. One-year EET status model’s predictors and respective odds ratios. Features selected by elastic net regularisation and respective odds ratios for one-year EET status model. EET = Employment, Education or Training; PANSS = Positive And Negative Symptoms Scale.

Fig 3. One-year point remission model’s predictors and respective odds ratios. Features selected by elastic net regularisation and respective odds ratios for one-year point remission model. PANSS = Positive And Negative Symptoms Scale.
treat, with limited evidence for benefit with conventional pharmacotherapy.[52] In contrast, there is better evidence in support of psychosocial interventions. For example, cognitive behavioural therapy has been shown to be effective up to 24-months.[53] Recent meta-analysis supports the use of interventions which enhance social cognition and interpersonal skills in the treatment of negative symptoms in psychosis.[54] Our finding highlights the requirement for a multimodal approach to treatment of FEP earlier in the illness course. Further, our model selects variables representative of higher social support including living with spouse and children and not renting as predictive of good symptomatic outcome which are consistent with established literature.[55]

Similarly, living with spouse and children is predictive of a good one-year EET outcome. Again, consistent with established literature, our model selects for good baseline EET and higher educational attainment as positive predictors of positive one-year EET outcome.[55] Interestingly, different individual PANSS items—P1 Delusions, P6 Suspiciousness and P7 Hostility—emerged as negative predictors of one-year EET. Such features may hinder service engagement and attachment which is known to influence recovery and adaption in FEP.[28] That distinct variables predict EET compared to those predicting remission, suggests the outcomes are capturing different yet complementary areas of recovery.

Our focus on EET status is timely and reflects the growing recognition within the literature that this functional outcome is at least as important as symptom remission. Young people not being in employment, education or training [NEET] has a considerable economic cost, accounting for a loss of €153bn or 1.2% of the European Union’s gross domestic product. Prolonged economic inactivity leads to serious mental health sequelae with higher rates of depression, alcohol or substance misuse and increased suicide attempts. Reducing youth unemployment remains a policy priority in high-income countries including the USA and Europe but the quality of evidence available to legislators surrounding NEET status is low.[56] We suggest that our robust finding related to the prediction of NEET outcome in first episode psychosis is the first high quality externally validated evidence.

The use of an externally (temporally and geographically) validated dataset with the same predictor and outcome variables is a major strength. Further, our models are interpretable, employ easily obtained variables and, especially in relation to prediction of EET status, are robust. However, our study has limitations. Small sample size relative to the number of

---

**Fig 4. One-year period remission model’s predictors and respective odds ratios.** Features selected by elastic net regularisation and respective odds ratios for one-year period remission model. PANSS = Positive And Negative Symptoms Scale.

https://doi.org/10.1371/journal.pone.0212846.g004
predictors is a potential concern. Plus, given its post-hoc nature, our analysis may have been underpowered. Using elastic net regularised regression, which simulation studies have shown outperforms conventional ordinary least squares or lasso while still enjoying the interpretability of feature selection missing with ridge, helps circumvent such concerns. Another concern may be that our cohorts have a statistically significant difference in accommodation status with more participants renting and less living in private accommodation with family in the training cohort as compared with the validation cohort. Similarly, changes in the wider macro-economic environment, such as those following the 2008 financial crisis which would have been felt by the training cohort only, are known to influence mental health and functional outcomes.[57] However, we would contend that such differences are in fact a strength of our study and reinforces the importance of external validation. Despite these issues, and that our model includes the accommodation status 'renting' as a negative predictor of EET outcome, our models still had a statistically significant and robust predictive performance in a geographically and temporally distinct cohort. However, we are not aware of published studies on how expert clinicians perform at predicting these outcomes which would be a required benchmark against which to compare new machine learning models. A further limitation is the lack of additional social variables, such as nicotine dependence[58], which observational studies have linked to outcomes in psychosis, or any cognitive and of biological markers of illness, including neuroimaging and blood markers, which previous studies have highlighted as important and may have improved predictive performance.[59]

These limitations should be addressed when designing future studies. A required next step prior to implementation into routine clinical practice would be to establish whether, by the accurate identification of individuals who will have poor outcomes, we can meaningful intervene to improve their prognosis. Further, few machine learning studies prospectively apply and update a prediction model to new patients in a clinical setting. Clinical problems are not stationary and intelligent algorithms should adapt to changing circumstance and response to treatment. Current research, including this present study, focuses on a classifier’s performance rather than its clinical usefulness. In a clinical context, often false negatives (under treatment) are more harmful than false positives (over treatment). Statistical performance should be weighted by the net benefit of potential harm resulting from either over or undertreatment versus the benefit of correct diagnosis, for example, by employing decision curve analysis.[27]

Conclusions

We have demonstrated that it is possible to accurately predict one-year symptomatic and functional employment education or training status in first episode psychosis. This has not been reported previously in an externally validated cohort. We propose that our results represent an important and exciting next step in unlocking the potential of predictive modelling in psychiatric illness.

Supporting information

S1 File. Supplemental methods.
(DOCX)

S2 File. R code for analysis.
(DOCX)

S3 File. Anonymised training and validation datasets.
(CSV)
Author Contributions

Conceptualization: Rajeev Krishnadas, Andrew I. Gumley.

Data curation: Kelly Chung, Alison Blair, Susie Brown, Suzy Clark, Kathryn Sowerbutts, Andrew I. Gumley.

Formal analysis: Samuel P. Leighton, Rajeev Krishnadas.

Funding acquisition: Alison Blair, Matthias Schwannauer, Andrew I. Gumley.

Methodology: Samuel P. Leighton, Rajeev Krishnadas, Matthias Schwannauer.

Project administration: Kelly Chung.

Supervision: Rajeev Krishnadas.

Validation: Rajeev Krishnadas.

Writing – original draft: Samuel P. Leighton, Rajeev Krishnadas.

Writing – review & editing: Matthias Schwannauer, Jonathan Cavanagh, Andrew I. Gumley.

References


34. Kuhn M. caret: Classification and Regression Training [Internet]. Available from: https://github.com/topepo/caret/


42. Seshan VE. clinfun: Clinical Trial Design and Data Analysis Functions [Internet]. 2017. Available from: https://cran.r-project.org/package=clinfun


