Development and Validation of Predictive Model for a Diagnosis of First Episode Psychosis Using the Multinational EU-GEI Case–control Study and Modern Statistical Learning Methods

Olesya Ajnakina1,2,8, Ihsan Fadilah1, Diego Quattrone3, Celso Arango4, Domenico Berardi5, Miguel Bernardo4, Julio Bobes7, Lieuwe de Haan8, Cristina Marta Del-Ben4, Charlotte Gayer-Anderson10, Simona Stilo11,12, Hannah E. Jongsma13,14, Antonio Lasalvia15,16, Sarah Tosato15,16, Pierre-Michel Llorca16, Paulo Rossi Menezes17, Bart P. Rutten18, Jose Luis Santos19, Julio Sanjuán20, Jean-Paul Selten21, Andrei Szöke22, Ilaria Tarricone23, Giuseppe D’Andrea5, Andrea Tortelli24, Eva Velthorst25,26, Peter B. Jones27,28, Manuel Arrojo Romero29, Caterina La Cascia30, James B. Kirkbride31,32,33, Michael O’Donovan34, Craig Morgan10, Marta di Forti11,12, EU-GEI WP2 Group§, and Daniel Stahl1

1Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, University of London, London, UK; 2Department of Behavioural Science and Health, Institute of Epidemiology and Health Care, University College London, London, UK; 3Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, London, UK; 4Child and Adolescent Psychiatry Department, Institute of Psychiatry and Mental Health, Hospital General Universitario Gregorio Marañón, School of Medicine, Universidade Complutense, IISGM, CIBERSAM, C/Doctor Esquerdo 46, 28007 Madrid, Spain; 5Department of Biomedical and Neuromotor Sciences, Psychiatry Unit, Alma Mater Studiorum Università di Bologna, Viale Pepoli 5, 40126 Bologna, Italy; 6Department of Psychiatry, Barcelona Clinic Schizophrenia Unit, Neuroscience Institute, Hospital Clinic of Barcelona, University of Barcelona, IDIBAPS, CIBERSAM, Barcelona, Spain; 7Faculty of Medicine and Health Sciences, Psychiatry, University of São Paulo, São Paulo, Brazil; 8Department of Health Service and Population Research, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, London, UK; 9Department of Mental Health and Addiction Services, ASP Crotone, Crotone, Italy; 10Department of Psychiatry Studies, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, London, UK; 11Centre for Transcultural Psychiatry Veldzicht, Balkbrug, The Netherlands; 12University Centre for Psychiatry, University Medical Centre Groningen, Groningen, The Netherlands; 13Section of Psychiatry, Department of Neuroscience, Biomedicine and Movement Sciences, University of Verona, Piazzale L.A. Scuro 10, 37134 Verona, Italy; 14Université Clermont Auvergne, CMP-B CHU, CNRS, Clermont Auvergne INP, Institut Pascal, F-63000 Clermont-Ferrand, France; 15Department of Preventative Medicine, Faculdeade de Medicina FMUSP, University of São Paulo, São Paulo, Brazil; 16Department of Psychiatry and Neuropsychiatry, School for Mental Health and Neuroscience, South Limburg Mental Health Research and Teaching Network, Maastricht University Medical Centre, P.O. Box 616, 6200 MD Maastricht, The Netherlands; 17Department of Psychiatry, Early Psychosis Section, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands; 18Department of Health and Service Population Research, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, London, UK; 19Department of Mental Health and Addiction Services, Asp Crotone, Crotone, Italy; 20Department of Psychiatry Studies, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, London, UK; 21Centre for Transcultural Psychiatry Veldzicht, Balkbrug, The Netherlands; 22University Centre for Psychiatry, University Medical Centre Groningen, Groningen, The Netherlands; 23Section of Psychiatry, Department of Neuroscience, Biomedicine and Movement Sciences, University of Verona, Piazzale L.A. Scuro 10, 37134 Verona, Italy; 24Université Clermont Auvergne, CMP-B CHU, CNRS, Clermont Auvergne INP, Institut Pascal, F-63000 Clermont-Ferrand, France; 25Department of Medical and Surgical Sciences, Bologna University, Bologna, Italy; 26Etablissement Public de Santé Maison Blanche, Paris, France; 27Department of Psychiatry, Early Psychosis Section, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands; 28Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 29Department of Psychiatry, University of Cambridge, Herchel Smith Building for Brain and Mind Sciences, Forvie Site, Robinson Way, Cambridge, CB2 0SZ, UK; 30CAMEO Early Intervention Service, Cambridgeshire and Peterborough NHS Foundation Trust, Cambridge, CB21 5EF, UK; 31Department of Psychiatry, Psychiatric Genetic Group, Instituto de Investigación Sanitaria de Santiago de Compostela, Complejo Hospitalario s, Santiago de Compostela, Spain; 32Department of Experimental Biomedicine and Clinical Neuroscience, University of Palermo, Via G. La Loggia 1, 90129 Palermo, Italy; 33Psylife Group, Division of Psychiatry, University College London, 6th Floor, Maple House, 149 Tottenham Court Road, London, W1T 7NF, UK; 34Department of Psychiatry, Brain Centre Rudolf Magnus, Utrecht University Medical centre, Utrecht, The Netherlands; 35Department of Psychiatry and Neuropsychiatry, School for Mental Health and Neuroscience, South Limburg Mental Health Research and Teaching Network, Maastricht University Medical Centre, P.O. Box 616, 6200 MD Maastricht, The Netherlands; 36Division of Psychological Medicine and Clinical Neurosciences, MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff CF24 4HQ, UK; 37Department of Psychiatry, Experimental Biomedicine and Clinical Neuroscience, University of Palermo, Palermo, Italy.

© The Author(s) 2023. Published by Oxford University Press on behalf of the University of Maryland’s school of medicine, Maryland Psychiatric Research Center.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
Introduction

First episode psychosis (FEP), which affects approximately 3% of the adult population, is an umbrella term used to refer to schizophrenia spectrum disorders or related psychotic disorders. Although schizophrenia was initially conceptualized as a chronic, progressive deteriorating condition, accumulating evidence suggests that people with a diagnosis of schizophrenia can experience symptomatic improvements and regain a degree of social and occupational functioning, especially when early intervention services intervene at the onset of the very first psychosis episode. This ignited an increased focus on specialist early intervention services for FEP, the aim of which is to reduce treatment delay, increase chance for recovery and improve overall prognosis of psychosis; however, the detection of those individuals who are at risk for developing FEP is currently limited. The reasons for the difficulties in detecting people who are at a greater risk for FEP are diverse including lack of financial recourses, high work-load and reliance on help-seeking behaviors. Indeed, most psychiatric services cannot offer a prompt assessment of the person at risk after the referral was made.

This recognition ignited development of individualized diagnostic prediction modeling for disease diagnosis that considers individual variability in characteristics and lifestyle of each person. Thus, diagnostic models, which having been built on combined effects of thoroughly selected predictors, can be used to forecast the probability of a certain condition being present at the individual level. This is particularly potent considering that detection of people who are at risk for FEP using a validated diagnostic model does not rely on help-seeking and can be implemented at differences services. For example, it has been shown that while 45% of the people who were unlimitedly diagnosed with FEP were referred to early intervention services by the emergency medical services with another 17.9% of the FEP patients came in contact with mental health services via the criminal justice agencies. These demonstrate that having a reliable tool to identify persons at risk of having FEP across these services will make the identification of people who are at risk for FEP much easier making the referrals to early interventions services more promptly. Therefore, it is hoped that availability of diagnostic models will facilitate a more rapid identification of individuals who are at a higher risk of FEP. This in turn would reduce time to treatment initiation, which is currently delayed for up to 3 years, subsequently minimizing the social and functional disability that results from prolonged untreated psychoses.

To-date, several studies, having employed either neuroimaging methods or proteomic data to relatively small samples, aimed to develop a diagnostic model to classify individuals with schizophrenia compared to healthy controls. However, the implementation of the models built on such complex data is likely to be constrained by logistical and financial challenges. Currently, there is no study that has developed, evaluated, and validated a diagnostic model for FEP using data reflecting the real-life
clinical information available to a physician and a patient at the time of the assessment. Imperatively, it remains unknown if a diagnostic model for classifying individual with FEP trained on data acquired at one site will perform similarly well on data acquired from a new site that was not included in the model development.19

Of course, this lack of progress in individualized diagnostic for FEP could be due to the complex etiology of FEP disorders, which may be intrinsically difficult to predict at an individual level. It is, nonetheless, equally feasible that the lack of progress can, at least in part, be attributed to significant methodological shortcomings that have engulfed the field of prediction modeling in psychiatry.14 These include relying on small samples, utilizing substantially lower numbers of cases relative to the number of considered predictor variables, employing unreliable methods to select predictor variables for inclusion into the model, not properly assessing the accuracy of the model, and not efficiently dealing with missing data.22–24 In fact, a recent systematic review showed that all current diagnostic models in FEP were at high risk of bias,14 making it highly unlikely for these models to be of any use.

In the era of precision medicine, computationally demanding modern statistical learning algorithms, particularly regularized regression methods (RRMs),25 promise to provide a useful tool for diagnostic modeling. Through an introduction of a penalty for overfitting, which occurs when the developed model provides an over-optimistic assessment of the predictive performance,23 RRM produce a model with good interpretability,26 which is especially portent for clinical application. Therefore, in the present study using a large multi-center phenotypically well-defined sample of FEP,27 we employed RRM to develop, evaluate, and validate a diagnostic risk estimation model to classify individual with FEP based on an individual profile of sociodemographic characteristics and environmental circumstances.28 The model was developed following the current guidelines.29,30 To ensure our model is appropriate for routine use in clinical practice,31 we used the internal-external validation in multi-site settings highlighting the extent to which the developed model can be generalized to the data from plausibly related settings.32

Methods

Study Design and Participants

Participants were recruited and assessed as part of the incidence and first episode case-control study, conducted as part of the EUPorean network of national schizophrenia networks investigating Gene-Environment Interactions (EU-GEI) study,27 which comprises the largest multi-site study of psychotic disorders ever conducted. EU-GEI study was established between May 2010 and April 2015 in tightly defined catchment areas in 17 sites across 6 countries, which were UK, The Netherlands, France, Spain, Italy and Brazil.33 The research sites within each country were purposefully selected to include a mix of urban and rural areas.27,33 All participants provided informed, written consent following full explanation of the study. It is noteworthy that the combined incidence and case-control methodology allowed us to account for any potential selection biases amongst the recruited and assessed cases.

Ethical Approval. All participants who agreed to take part in the case–control study provided informed, written consent following full explanation of the study. Ethical approval for the study was provided by relevant local research ethics committees in each of the study sites.27,33

Ascertainment of Cases. The inclusion criteria for FEP cases were: (1) presentation with a clinical diagnosis for an untreated FEP as defined by International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10) criteria22 (codes F20–F33) within the timeframe of the study; (2) aged between 18 and 64 years (inclusive); and (3) resident within one of the 17 defined catchment area at the time of their first presentation to psychiatric services for psychosis. Exclusion criteria were: (1) a previous contact with specialist mental health services for psychotic symptoms outside of the study period at each site; (2) evidence of psychotic symptoms precipitated by an organic cause (ICD-10: F09); (3) transient psychotic symptoms resulting from acute intoxication (F1.x.5); (4) severe learning disabilities, defined by an IQ less than 50 or diagnosis of intellectual disability (F70–F79); and (5) insufficient fluency of the primary language at each site to complete assessments.27

Ascertainment of Controls. To better reflect the source population from which the cases arose, controls were recruited based on random and quota sampling that considered the distribution of age, sex, and ethnicity in each region. Inclusion criteria for controls were: (1) age 18–64 years; (2) resident in the distinctly defined catchment region; (3) adequate fluency of the primary language used in each site; (4) no history of current or past psychiatric disorders.27,33 The individuals who were recruited as controls for the study were broadly representative of local populations in relation to age, gender, and ethnicity.27 Individuals who agreed to take part were screened for a history of psychosis. Those who reported previous or current treatment for psychosis were excluded.27

Predictors. Following a previous research protocol,34 we excluded variables which had a high collinearity with other variables, and/or had > 50% missing values. Overall, 96 predictors related to participants’ sociodemographic circumstances, childhood adversity, life events experienced in adulthood and substance use were included in the model development (Supplementary table 1).
Information on these predictors was collected using previously validated tools with a good inter-rater reliability and structured, standardized format across sites.\textsuperscript{27,33,35}

Statistical Analysis

The process of model development, evaluation and validation was carried out according to methodological standards outlined by Steyerberg et al\textsuperscript{25}, results were reported according to the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines\textsuperscript{36}; the completed checklist is provided in Supplementary table 2. All analyses were performed in RStudio release 4.02.\textsuperscript{37}

Sample Size Calculations. To calculate the sample size in the present study, we utilized the guidelines for sample size calculations when developing prediction models for binary outcomes.\textsuperscript{38} These guidelines require to account for not only on the number of events relative to the number of candidate predictor, which is a well-known rule of thumb for the required sample size,\textsuperscript{39} but also on the total number of participants, the outcome proportion (incidence) in the study population, and the expected predictive performance of the model. These information is used to tailor sample size requirements to the specific setting of interest, with the aim of minimizing the potential for model overfitting while targeting precise estimates of key parameters.\textsuperscript{38} Accordingly, to assess if our sample size was large enough to develop a robust prediction diagnostic model, we calculated the required sample size according to these guidelines\textsuperscript{38} considering several parameters (Supplementary Material). Assuming the value of $R^2$ corresponds to an $R^2_{\text{Nagelkerke}}$ of 0.15, that is, $R^2_{\text{CS}} = 0.15 \times \max([R^2_{\text{CS}}])$,\textsuperscript{38} the sample size required for the model development was $n = 2816$ corresponding to 14.67 events per predictor. Consequently, our sample size of $n = 2627$ was slightly below the requirement; though, this sample size calculation did not consider regularization, which reduces the risk of overfitting.

Imputation of Missing Values. In the present study, some variables had missing values (Supplementary table 4) with an average missingness in the entire sample was approximately 12.9%. To avoid using unrepresentative sample of complete cases that may result in incorrect risk predictions,\textsuperscript{40,41} we imputed missing values employing missForest, which is an imputation method based on random forest that handles continuous and categorical variables equally well and accommodates non-linear relation structures.\textsuperscript{42,43} As recommended for prediction models, the outcome was included in the imputation process.\textsuperscript{44} The distribution of the variables included in the analyses before and after the imputation are presented in Supplementary table 3 showing that the imputed values were very closely aligned with the observed values across all variables used in the analyses.

Variable Selection and Model Fitting. To build the diagnostic model and identify which of included predictors were important for classifying individual with FEP and controls, we applied a binary logistic model with regularization by the least absolute shrinkage and selection operator (LASSO).\textsuperscript{45} LASSO entails fitting a model, which, by imposing penalty ($\lambda$) on the size of regression parameter estimates shrink them towards zero,\textsuperscript{46,47} simultaneously selects predictors, estimates their effects, and introduces parsimony. Therefore, if a suitable $\lambda$ is chosen, LASSO intrinsically performs predictor selection and deals with collinearity. Selection of the optimal tuning parameter $\lambda$ optimizing the model performance is described below.

Model Estimation. The tuning parameter $\lambda$ optimizing the area under the receiver operating characteristic curve (AUROC) was chosen from a grid of 100 $\lambda$ values through 10-fold repeated cross-validation (CV).\textsuperscript{45} 10-fold CV divided data randomly into 10 non-overlapping data partitions; participants included in the first 9 partitions were considered as the training sample, and the remaining individuals as the test sample. To reduce the variance of CV, 10-fold CV was repeated 10 times computing the AUROC for each $\lambda$ value. As a parsimonious model is desirable for practice\textsuperscript{46} and may generalize better to different populations,\textsuperscript{49} though often at the expense of a lower predictive performance, the model that corresponded to 3% tolerance of the maximum AUROC yielding more parsimony with fewer irrelevant variables compared to standard minimum lambda,\textsuperscript{46} was chosen as the final model.

Model Performance. Model’s accuracy was measured with discrimination and calibration. Discrimination indicates how well a model separates individuals who experienced an event from those who did not; we assessed discrimination using AUROC, where a value of 0.5 indicates that a model does not discriminate better than chance, while 1 indicates that a model discriminates perfectly.\textsuperscript{50} Calibration, assessed via calibration slope $\beta$, which should ideally be 1, and the calibration-in-the-large $\alpha$, which ideally should be zero, describes how well the predicted risk corresponds to the risk from the observed data\textsuperscript{31,32} and can be described as a measure of bias in a model.\textsuperscript{53} We present calibration graphically by placing the estimated and actual outcome risk on the horizontal and vertical axes, respectively. We further measured the prediction accuracy of our models with sensitivity and specificity. Unlike the traditional 50%, which follows often incorrect assumption that the false-positive and false-negative are equally important,\textsuperscript{25} to classify an individual as high or low risk based on a prediction model, a cut-off for the predicted probability (ie, “decision threshold”)\textsuperscript{25} was selected by maximizing the sum of the model’s
sensitivity and specificity to minimize the false positives, which are unavoidable. This entailed selecting the decision threshold that maximized the overall correct classification rates, while choosing the point on the receiver operating characteristic curve farthest from chance.

Model Validation. We validated the developed model using internal-external cross-country validation, which allows quantifying the generalizability of a prediction model across different settings and populations. Specifically, having six countries, we grouped participants by country and redeveloped the model, repeating every step of estimating and selecting candidate predictors, in five of the six countries. We then evaluated the resulting model using the data from the remaining country measuring discrimination and calibration as described above. We repeated this validation algorithm six times until each country was used as a validation sample and reported the mean as general performance estimates. The full model equation is presented in Supplementary materials to enable the essential independent external validation.

Results

Study Participants

The sample characteristics are provided in table 1. The sample comprised 2627 participants; of these, 43.0% (n = 1130) had a diagnosis of FEP and 57.9% (n = 1497) were participants without FEP disorders. The average age of participants with a diagnosis of FEP at the time of the assessment was 31.3 years (SD = 10.6), 61.7% (n = 697) were men and 63.3% (n = 715) were of white ethnicity. Compared to FEP group, participants without FEP were older (mean = 36.3 years, SD = 12.9) of whom 47.2% (n = 706) were men and 78.7% (n = 1178) were of white ethnicity. Of the entire sample, 22.2% (n = 582) were recruited from the UK, 15.5% (n = 406) were recruited from The Netherlands, 16.2% (n = 426) came from Spain, 9.6% (n = 252) came from France, 17.8% (n = 467) were recruited from Italy, and 18.8% (n = 494) came from Brazil. The participants recruited across all countries were comparable in terms of age, gender, and relationship status at the time of the assessment (table 2).

Diagnostic Model

Internally-externally validated performance of our model is presented in table 3. The model included 22 (22.9% out of n = 96) predictor variables (Supplementary table 4). The model’s apparent performance is presented in Supplementary table 5. Following model’s validation, a very good discrimination was observed across all countries (range AUROC = 0.84–0.86). The calibration intercept (α) for all, but The Netherlands (calibration intercept [α] = 0.56), countries was slightly larger than 0 (range = −0.24 to 0.13). Calibration slope (β) was the lowest for the sample obtained in Brazil (calibration slope [β] = 1.11) and Spain (calibration slope [β] = 1.22) indicating an excellent

Table 1. Characteristics of Participants in the Study Population

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Overall n = 2627</th>
<th>FEP n = 1130 (43.0%)</th>
<th>Control n = 1497 (57.9%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at assessment (years)</td>
<td>34.0 ± 12.2</td>
<td>31.3 ± 10.6</td>
<td>36.1 ± 12.9</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1403 ± 53.4</td>
<td>697 ± 61.7</td>
<td>706 ± 47.2</td>
</tr>
<tr>
<td>Female</td>
<td>1224 ± 46.6</td>
<td>433 ± 38.3</td>
<td>791 ± 52.8</td>
</tr>
<tr>
<td>Education (years)</td>
<td>14.0 ± 4.3</td>
<td>12.9 ± 4.2</td>
<td>14.7 ± 4.2</td>
</tr>
<tr>
<td>Ever been employed</td>
<td>2402 ± 91.6</td>
<td>999 ± 88.6</td>
<td>1403 ± 93.8</td>
</tr>
<tr>
<td>Never been in a long-term relationship</td>
<td>507 ± 19.3</td>
<td>345 ± 30.5</td>
<td>162 ± 10.8</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1893 ± 72.1</td>
<td>715 ± 63.3</td>
<td>1178 ± 78.7</td>
</tr>
<tr>
<td>Black</td>
<td>380 ± 14.5</td>
<td>235 ± 20.8</td>
<td>145 ± 9.7</td>
</tr>
<tr>
<td>Other</td>
<td>353 ± 13.4</td>
<td>180 ± 15.9</td>
<td>173 ± 11.6</td>
</tr>
<tr>
<td>Country</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The UK</td>
<td>582 ± 22.2</td>
<td>246 ± 21.8</td>
<td>336 ± 22.4</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>406 ± 15.5</td>
<td>196 ± 17.3</td>
<td>210 ± 14.0</td>
</tr>
<tr>
<td>Spain</td>
<td>426 ± 16.2</td>
<td>204 ± 18.1</td>
<td>222 ± 14.8</td>
</tr>
<tr>
<td>France</td>
<td>252 ± 9.6</td>
<td>105 ± 9.3</td>
<td>147 ± 9.8</td>
</tr>
<tr>
<td>Italy</td>
<td>467 ± 17.8</td>
<td>187 ± 16.5</td>
<td>280 ± 18.7</td>
</tr>
<tr>
<td>Brazil</td>
<td>494 ± 18.8</td>
<td>192 ± 17.0</td>
<td>302 ± 20.2</td>
</tr>
</tbody>
</table>

Note: sd, standard deviation.
model prediction accuracy; though, calibration slope ($\beta$) for France, the UK and The Netherlands were relatively high (1.77, 1.71, and 1.61, respectively) suggesting that the model for these sites slightly overestimate risks. Calibration plots showed good agreement between observed and expected risk at predicted probabilities across all countries (figure 1). Using a cut-off point of 39.8%, our model was able to discriminate adults FEP from controls with good sensitivity (range = 73.9–78.0%) and specificity (range = 75.6–79.3%).

**Discussion**

Having utilized data across 17 research sites from 6 countries to our knowledge, this is the first study to develop, fully evaluate and validate a diagnostic model for classifying FEP based on a personal profile of 22 personal characteristics and lifestyle. We followed the current guidelines for model development, evaluation and validation. Our results indicate that classification of FEP and controls is possible with high predictive accuracy across the UK, The Netherlands, Spain France Italy and even Brazil. To maximize the predictive accuracy, we catered for incomplete data, which is a common but serious limitation in psychiatric research but generally not addressed sufficiently. Given that the data ascertainment for this study was carried out in major urban and rural sites with heterogeneous populations suggests that the validity of our model may extend to other centers with similar population profiles. Sensitivity and specificity are tests of accuracy of a model and are among the fundamental measures to understanding the utility of clinical tests. Sensitivity refers to how good a test is at correctly identifying people who have the disease; whereas the specificity of a clinical test refers to the ability of the test to correctly identify those patients without the disease. Our results demonstrate that our model has a high sensitivity and high specificity implying that it will detect accurately many adults who are disease free as low risk without recurring further investigation. Because the model does not require any laboratory testing or clinical measurements, it could be easily integrated into electronic case-registers to facilitate the automatic and individualized diagnostic

---

**Table 2. Characteristics of Participants in the Study Population, by Country**

<table>
<thead>
<tr>
<th>Recruitment countries</th>
<th>The UK $n = 582$ (22.2%)</th>
<th>The Netherlands $n = 406$ (15.5%)</th>
<th>Spain $n = 426$ (16.2%)</th>
<th>France $n = 252$ (9.6%)</th>
<th>Italy $n = 467$ (17.8%)</th>
<th>Brazil $n = 494$ (18.8%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at assessment (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>31 17</td>
<td>31 22</td>
<td>34 18</td>
<td>33 22</td>
<td>31 19</td>
<td>30 14</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>315 54.1</td>
<td>235 57.9</td>
<td>188 44.1</td>
<td>134 53.2</td>
<td>228 48.8</td>
<td>253 51.2</td>
</tr>
<tr>
<td>Female</td>
<td>267 45.9</td>
<td>171 42.1</td>
<td>238 55.9</td>
<td>118 46.8</td>
<td>239 46.8</td>
<td>241 48.8</td>
</tr>
<tr>
<td>Education (years)</td>
<td>15.3 3.4</td>
<td>16.3 3.8</td>
<td>13.4 4.5</td>
<td>13.3 3.7</td>
<td>14.0 3.9</td>
<td>11.2 4.3</td>
</tr>
<tr>
<td>Ever been employed</td>
<td>539 92.6</td>
<td>398 98.0</td>
<td>385 90.8</td>
<td>229 91.2</td>
<td>396 85.2</td>
<td>455 92.1</td>
</tr>
<tr>
<td>Never been in a long-term relationship</td>
<td>110 18.9</td>
<td>94 23.2</td>
<td>91 21.4</td>
<td>35 13.9</td>
<td>89 19.1</td>
<td>88 17.8</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>348 59.8</td>
<td>292 72.1</td>
<td>382 89.7</td>
<td>143 56.7</td>
<td>434 92.9</td>
<td>294 59.5</td>
</tr>
<tr>
<td>Black</td>
<td>170 29.2</td>
<td>57 14.1</td>
<td>9 2.1</td>
<td>81 32.1</td>
<td>12 2.6</td>
<td>51 10.3</td>
</tr>
<tr>
<td>Other</td>
<td>64 11.0</td>
<td>56 13.8</td>
<td>35 8.2</td>
<td>28 11.1</td>
<td>21 4.5</td>
<td>149 30.2</td>
</tr>
</tbody>
</table>

Note: FEP, first-episode psychosis; N, number of participants; IQR, interquartile range; df, degrees of freedom.

**Predictor Variables.** Several predictor variables selected in the model (Supplementary table 4), such as educational attainment (unstandardized $\beta = 0.035$), being foreign-born (unstandardized $\beta = 0.113$), and childhood experiences including not having peers to go to (unstandardized $\beta = 0.264$), experiences of prolonged loneliness (unstandardized $\beta = 0.152$) and running away from home (unstandardized $\beta = 0.052$) existed before FEP onset. Thus, these variables may be seen as potentially causative predictors for developing FEP. Other important contributing factors for diagnostic risk for FEP were being unemployed (unstandardized $\beta = 0.671$), being single (unstandardized $\beta = 0.577$), having problems with the police (unstandardized $\beta = 0.484$) or having difficulties at work (unstandardized $\beta = 0.528$), using more cannabis than intended (unstandardized $\beta = 0.523$), daily cigarette smoking (unstandardized $\beta = 0.590$), and using other substances, such as cocaine (unstandardized $\beta = 0.227$). A worked example of calculating an individualized risk for FEP is provided in Supplementary Material.
identification of FEP based on electronic or clinical records.

The predictor variables that were selected by our model, such as lower level of educational attainment, childhood adversity and stressful life events in adulthood, and cannabis use have previously been linked to FEP risk and probably occurred before the onset of psychoses. Accordingly, our results reiterate the important role these experiences play in increasing risk for FEP providing avenues for prevention strategies. Importantly, some of these factors, such as cannabis use, childhood adversities and educational attainment are potentially preventable with the right interventions. For example, using the same data as in the present study, it was shown that 24% of FEP cases would have been prevented if none in the population consumed cannabis of high potency. Our results further reiterate that better educational attainment may protect from FEP risk perhaps via more effective coping strategies, healthier behaviors and social relationships. For many, however, FEP develops during a period critical to the consolidation of life skills, which may result in an individual being unable to obtain qualifications after illness onset. It is, therefore, imperative to provide people with FEP access to supported education programs to (re)-engage them in the workforce. The confirmation of these factors as pivotal in development of FEP further supports the long-term benefits of reducing an exposure to these risk factors for psychosis; though this likely to be challenging considering the pathogenic mechanism underlying the link between some of these risk factors and psychosis is not fully understood. Furthermore, it may be very difficult to diminish exposure to some risk factors, for example, child abuse or migration; though, an obvious place to start is by attempting to reduce society’s consumption of high-potency cannabis through public education.

Nonetheless, because all individuals with a diagnosis of FEP in the present study were already under the care of mental health services upon recruitment, it may be argued that there is a window of missed opportunity for detection of FEP before the illness onset using this model. An alternative approach is to develop a prognostic model that will aim to estimate an individual risk for FEP onset among young help-seeking people who have been identified as at clinical high risk, which is a state characterized by either “attenuated” psychotic symptoms, or full-blown psychotic symptoms that are brief and self-limiting. While these prognostic risk estimation studies offer a promise for detecting young people who are at high risk for converting to FEP from experiencing suboptimal symptoms, those young adults who have been classified as “at clinical high risk” are not representative of those who develop FEP in terms of socio-economic status, life-experiences and ethnical composition. In contract, our model was developed specifically for true
Fig. 1. Internally–externally validated calibration plots for the prediction model of FEP in the United Kingdom, the Netherlands, France, Spain, Italy, and Brazil. Squares illustrate risk groups by fourths of equally spaced model-estimated risks, through which the linear line was fitted (red). The smoothed loess curve (orange) was fitted based on individual data points. The 45° line (gray) represents perfect calibration where model-estimated equal actual risk. The histogram on the upper margin represents the distribution of model-estimated risks.
FEP cases, which ensures its generalizability to the wider FEP communities.

Calibration slope for our prediction diagnostic model, however, showed a presence of some overfitting when applied specifically to participants from France, the UK, and The Netherlands suggesting that for the patients recruited in these countries estimated risks may be too high for those who are at high risk and too low for those who are at low risk; though, the observed calibration slope estimates were still within a range of previously reported models. Nonetheless, there is no evidence to suggest that more complex models or encompassing biomarkers lead to significant improvements in prediction accuracy even at the expense of reduced interpretability compared with simpler statistical models.

The present study is not without limitations. Because we developed the model on a case-control sample, the true prevalence of FEP in the general population differs considerably. Although it may be argued that the percentage of missing values across variables might have affected the imputations and induced some bias in the estimates of the effects in the model, the proportion of missingness in the present study was comparable to many longitudinal datasets and within the range for *missForest* to handle it efficiently. As with many risk models, we only accounted for baseline variables, although for many time-varying factors, exposure status may change over time. However, using baseline variables reflects the real-life clinical information available to a physician and a participant when they need to make decisions on the likely risk of developing FEP disorders. Even though the average age between participants without a diagnosis of FEP was older than FEP participants, the age of onset of FEP in our sample was consistent with many studies on FEP conducted across Europe, and other continents. A higher proportion of our participants with FEP were from ethnic minorities when compared to health controls. However, psychiatric epidemiology has consistently demonstrated that the incidence rates of psychotic disorders are considerably elevated among those of Black ethnicity residing in the UK compared to the host population. Therefore, it is expected that individuals with FEP will be different greatly in ethnicity compared to adults without FEP diagnosis. It may be argued that many diagnostic categories assigned to patients on first contact with mental health services may either be provisional or likely to change over the illness course. Nonetheless, in the present study we focused on the baseline diagnosis to emulate the naturalistic setting for all patients with FEP when predicting their onset. There is further robust meta-analytical evidence for high prospective diagnostic stability in schizophrenia spectrum and affective spectrum psychoses in the due course of the illnesses. Nonetheless, it may be feasible to assume that there may be different predictors for affective versus non-affective psychosis. Thus, further modeling approaches may be necessary to investigate this in the possibility future. Finally, in the present study we have developed the model for FEP rather than individual diagnoses, such bipolar disorder, depression with psychotic features, because many diagnostic categories assigned to patients on first contact with mental health services may either be provisional or likely to change over the illness course.

**Conclusions**

Having employed modern statistical learning algorithms, we developed, evaluated, and validated a diagnostic model for classifying FEP that achieved a good discrimination and calibration across six European countries and Brazil supporting its robustness and good generalizability across FEP programs in different countries. This study, therefore, bears important implications for the development of affordable and easy-to-administer standardized assessment batteries that can evaluate individuals’ risk for FEP in clinical settings across countries with similar characteristics of adults with FEP.

**Supplementary Material**

Supplementary data are available at Schizophrenia Bulletin Open online.

**EU-GEI collaborators**

Kathryn Hubbard, Stephanie Beards, Doriana Cristofalo, Mara Parelleda, Pedro Cuadrado, José Juan Rodriguez Solano, David Fraguas, Álvaro Andreu-Bernabeu, Angel Carracedo, Enrique García Bernardo, Laura Roldán, Gonzalo López, Silvia Amoretti, Juan Nacher, Paz García-Portilla, Javier Costas, Estela Jiménez-López, Mario Mattei, Marta Rapado Castro, Emilio González, Covadonga Martínez, Emilio Sánchez, Manuel Durán-Cutilla, Nathalie Franke, Fabian Termorshuizen, Daniella van Dam, Elsje van der Ven, Elles Messchaart, Marion Leboyer, Franck Schürhoff, Stéphane Jamain, Grégoire Baudin, Aziz Ferchiou, Baptiste Pignon, Jean-Romain Richard, Thomas Charpeaud, Anne-Marie Tronche, Flora Frijda, Daniele La Barbera, Giovanna Marrazzo, Lucia Sideli, Crocettarachele Sartorio, Laura Ferraro, Fabio Seminiero, Camila Marcelino Loureiro, Rosana Shuhama, Mirella Ruggeri, Antonio LaSalvia, Chiara Bonetto.

1Department of Health Service and Population Research, Institute of Psychiatry, King’s College London,
De Crespigny Park, Denmark Hill, London, SE5 8AF, United Kingdom; 2Section of Psychiatry, Department of Neuroscience, Biomedicine and Movement, University of Verona, Piazzale L.A. Scuro 10, 37134 Verona, Italy; 3Department of Child and Adolescent Psychiatry, Hospital General Universitario Gregorio Marañón, School of Medicine, Universidad Complutense, IISGVM (CIBERSAM), C/Doctor Esquerdo 46, 28007 Madrid, Spain; 4Villa de Vallecas Mental Health Department, Hospital Universitario Infanta Leonor/ Hospital Virgen de la Torre, C/San Claudio 154, 28038 Madrid, Spain; 5Puente de Vallecas Mental Health Department, Hospital Universitario Infanta Leonor/ Hospital Virgen de la Torre, C/San de Vallecas Mental Health Centre, Hospital Universitario Infanta Leonor/ Hospital Virgen de la Torre, C/San Claudio 154, 28038 Madrid, Spain; 6Equipe 15, 51 Avenue de Maréchal de Lattre de Tassigny, 94010 Créteil, France; 7Department of Psychiatry, Hospital General Universitario Gregorio Marañón, School of Medicine, Universidad Complutense, IISGVM (CIBERSAM), C/Doctor Esquerdo 46, 28007 Madrid, Spain; 8Department of Psychiatry, Hospital Universitario Infanta Leonor/ Hospital Virgen de la Torre, Centro de Salud Mental Puente de Vallecas, C/Peña Gorbea 4, 28018 Madrid, Spain; 9Fundación Pública Galega de Medicina Xenómica, Hospital Clínico Universitario, Choupana s/n, 15782 Santiago de Compostela, Spain; 10Department of Psychiatry, Hospital General Universitario Gregorio Marañón, School of Medicine, Universidad Complutense, IISGVM (CIBERSAM), C/Doctor Esquerdo 46, 28007 Madrid, Spain; 11Department of Psychiatry, Hospital Clínico San Carlos, Madrid, Spain; 12Department of Psychiatry, University of Palermo, Via G. La Loggia n.1, 90129 Palermo, Italy; 13Unit of Psychiatry, “P. Giacone” General Hospital, Via G. La Loggia n.1, 90129 Palermo, Italy; 14Department of Neurociências e Ciências do Comportamento, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Av. Bandeirantes, 3900 - Monte Alegre-CEP 14049-900, Ribeirão Preto, SP, Brasil; 15Núcleo de Pesquisa em Saúde Mental Populacional, Universidade de São Paulo, Avenida Doutor Arnaldo 455, CEP 01246-903, SP, Brasil

Funding

OA is funded by the National Institute for Health Research (NIHR) (NIHR Post-Doctoral Fellowship—PDF-2018-11-ST2-020). IF is funded by NIHR Predoctoral Fellowship (NIHR300493). MDF is funded by Clinician Scientist Medical Research Council fellowship (project reference MR/M008436/1). DQ is funded by Post-Doctoral Guarantors of Brain Clinical Fellowship. DS is funded part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King’s College London. JBK is supported by National Institute for Health Research, University College London Hospital, Biomedical Research Centre. The EU-GEI Project is funded by the European Community’s Seventh Framework Programme under grant agreement No. HEALTH-F2-2010-241909 (Project EU-GEI). The Brazilian study was funded by the São Paulo Research Foundation under grant number 2012/0417-0. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health and Social Care.

Conflict of Interest

R.M.M. has received honoraria from Janssen, Sunovian, Lundbeck and Otsuka. M.B. has been a consultant for, received grant/research support and honoraria from, and been on the speakers/advisory board of ABBiotics, Adamed, Angelini, Casen Recordati, Janssen-Cilag, Menarini, Rovi and Takeda. Other authors declare that they have no conflict of interest. All other authors declare no conflict of interest.


