



**HAL**  
open science

## **Prediction of relapse in a French cohort of outpatients with schizophrenia (FACE-SZ): Prediction, not association.**

Susana Barbosa, Ryad Tamouza, Marion Leboyer, Bruno Aouizerate, Christelle Andrieu, Myrtille Andre, Wahid Boukouaci, Delphine Capdevielle, Isabelle Chereau, Julie Clauss Kobayashi, et al.

### ► **To cite this version:**

Susana Barbosa, Ryad Tamouza, Marion Leboyer, Bruno Aouizerate, Christelle Andrieu, et al.. Prediction of relapse in a French cohort of outpatients with schizophrenia (FACE-SZ): Prediction, not association.. Progress in Neuro-Psychopharmacology and Biological Psychiatry, 2025, 137, <10.1016/j.pnpbp.2025.111304>. <hal-05029188>

**HAL Id: hal-05029188**

**<https://hal.science/hal-05029188v1>**

Submitted on 25 Apr 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

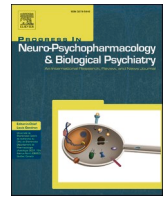
L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

# Progress in Neuropsychopharmacology & Biological Psychiatry

journal homepage: [www.elsevier.com/locate/pnp](http://www.elsevier.com/locate/pnp)

## Prediction of relapse in a French cohort of outpatients with schizophrenia (FACE-SZ): Prediction, not association.

Susana Barbosa<sup>a</sup>, Ryad Tamouza<sup>b,c,d</sup>, Marion Leboyer<sup>b,c,d</sup>, Bruno Auizerate<sup>b,e,f</sup>,  
Christelle Andrieu<sup>b,g</sup>, Myrtille Andre<sup>b,h</sup>, Wahid Boukouaci<sup>b,c,d</sup>, Delphine Capdevielle<sup>b,h</sup>,  
Isabelle Chereau<sup>b,i</sup>, Julie Clauss Kobayashi<sup>b,j</sup>, Nathalie Coulon<sup>b,k</sup>, Jean-Michel Dorey<sup>b,l</sup>,  
Laetitia Davidovic<sup>a</sup>, Caroline Dubertret<sup>b,m,n</sup>, Eric Fakra<sup>b,o</sup>, Guillaume Fond<sup>b,g</sup>, Tudi Goze<sup>b,p</sup>,  
Olfa Khalfallah<sup>a</sup>, Sylvain Leignier<sup>b,k</sup>, Pierre Michel Llorca<sup>b,i</sup>, Jasmina Mallet<sup>b,m,n</sup>,  
Emanuela Martinuzzi<sup>a</sup>, David Misdrahi<sup>b,q,r</sup>, Nicolas Oriol<sup>b,o</sup>, Baptiste Pignon<sup>b,c,d</sup>,  
Romain Rey<sup>b,l</sup>, Paul Roux<sup>b,s,t</sup>, Franck Schürhoff<sup>b,c,d</sup>, Benoit Schorr<sup>b,j</sup>, Mathieu Urbach<sup>b,s,t</sup>,  
Etienne Very<sup>b,p</sup>, Ching-Lien Wu<sup>b,c,d</sup>, Michael Benros<sup>u</sup>, Judit Simon<sup>v,w</sup>, Alkomiet Hasan<sup>x,y</sup>, for  
The FondaMental Academic Centers of Expertise for Schizophrenia (FACE-SZ) collaborators,  
Nicolas Glaichenhaus<sup>a,b</sup>, Ophélie Godin<sup>b,c,\*</sup>

<sup>a</sup> Université Côte d'Azur, Centre National de la Recherche Scientifique, Institut de Pharmacologie Moléculaire et Cellulaire, Valbonne, France

<sup>b</sup> Fondation FondaMental, Créteil, France

<sup>c</sup> Université Paris Est Créteil (UPEC), INSERM, Institut Henri Mondor the Recherche Biomédicale, Laboratoire Neuro-Psychiatrie Translationnelle, Créteil, France

<sup>d</sup> Assistance Publique Hôpitaux de Paris, Hôpitaux Universitaires Henri Mondor, Département Médico-Universitaire de Psychiatrie et d'Addictologie, Créteil, France

<sup>e</sup> University Department of General Psychiatry, Charles Perrens Hospital, F-33076 Bordeaux, France

<sup>f</sup> Laboratory of Nutrition and Integrative Neurobiology (UMR INRA 1286), University of Bordeaux, Bordeaux, France

<sup>g</sup> AP-HM, Service Universitaire de psychiatrie et d'addictologie (Pr Lançon), Hôpital Sainte Marguerite, 13009 Marseille, France

<sup>h</sup> Service Universitaire de Psychiatrie Adulte, Hôpital la Colombière, CHRU Montpellier, Université Montpellier 1, Inserm, 1061 Montpellier, France

<sup>i</sup> University of Clermont Auvergne, CMP-B CHU, CNRS, Clermont Auvergne INP, Institut Pascal, F-63000 Clermont-Ferrand, France

<sup>j</sup> Hôpitaux Universitaires de Strasbourg, Université de Strasbourg, INSERM U1329, Fédération de Médecine Translationnelle de Strasbourg, Strasbourg, France

<sup>k</sup> Centre Référent de Réhabilitation Psychosociale et de Remédiation Cognitive (C3R), CH Alpes Isère, Grenoble, France

<sup>l</sup> INSERM U1028, CNRS UMR5292, Centre de Recherche en Neurosciences de Lyon, Université Claude Bernard Lyon 1, Equipe PSYR2, Centre Hospitalier Le Vinatier, Pole Est, 95 bd Pinel, BP 30039, 69678 Bron Cedex, France

<sup>m</sup> AP-HP, Groupe Hospitalo-universitaire AP-HP Nord, DMU ESPRIT, Service de Psychiatrie et Addictologie, Hôpital Louis Mourier, Colombes, France

<sup>n</sup> Université de Paris Inserm UMR1266, Sorbonne Paris Cité, Faculté de Médecine, Paris, France

<sup>o</sup> Department of Psychiatry, University Hospital of Saint-Etienne, Saint-Etienne, France

<sup>p</sup> Department of Psychiatry, Psychotherapies, Art-therapy, Toulouse University Hospital; Equipe de Recherche sur les Rationalités Philosophiques et les Savoirs - EA3051, Université de Toulouse - Jean Jaurès, Toulouse, France

<sup>q</sup> University of Bordeaux, Aquitaine Institute for Cognitive and Integrative Neuroscience (CNRS UMR 5287-INCIA), Bordeaux, France

<sup>r</sup> Department of Adult Psychiatry, Charles Perrens Hospital, Bordeaux, France

<sup>s</sup> Versailles Hospital, Department of Adult Psychiatry and Addictology, Centre Hospitalier de Versailles, 177 rue de Versailles, 78157 Le Chesnay, France

<sup>t</sup> DisAP-DevPsy-CESP, INSERM UMR1018, University of Paris-Saclay; University of Versailles Saint-Quentin-En-Yvelines, 78000 Versailles, France

<sup>u</sup> CORE-Copenhagen Research Centre for Mental Health, Mental Health Centre Copenhagen, Copenhagen University Hospital, Copenhagen, Denmark

<sup>v</sup> Department of Health Economics, Center for Public Health, Medical University of Vienna, Vienna, Austria

<sup>w</sup> Department of Psychiatry, University of Oxford, Oxford, UK

<sup>x</sup> Department of Psychiatry, Psychotherapy, and Psychosomatics, Medical Faculty, University of Augsburg, Bezirkskrankenhaus Augsburg, Geschwister-Schönert-Str. 1, 86156 Augsburg, Germany

<sup>y</sup> DZPG (German Center for Mental Health), partner site München/Augsburg, Augsburg, Germany

\* Corresponding author at: Inserm U955 – IMRB (Eq Translational Neuropsychiatry), Fondation FondaMental, Hôpital Albert Chenevier - Pôle Psychiatrie - 40 rue de Mesly, 94000 Créteil, France.

E-mail address: [ophelia.godin@inserm.fr](mailto:ophelia.godin@inserm.fr) (O. Godin).

<https://doi.org/10.1016/j.pnpbp.2025.111304>

Received 29 October 2024; Received in revised form 26 February 2025; Accepted 26 February 2025

Available online 27 February 2025

0278-5846/© 2025 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## ARTICLE INFO

## Keywords:

Schizophrenia  
Relapse  
Prediction  
Machine learning  
Inflammation  
Immune system  
Word count = 3637.

## ABSTRACT

**Background:** Schizophrenia (SZ) commonly manifests through multiple relapses, each impeding the path to recovery and incurring personal and societal costs. Despite the identification of various risk factors associated to the risk of relapse, the development of accurate algorithms predictive of relapse has been limited, partly due to inadequate statistical methods. Additionally, despite the wealth of data showing strong associations between inflammation and schizophrenia, the two existing studies failed to demonstrate whether inflammatory parameters could predict relapse. Our goal is then to identify clinical and inflammatory parameters associated with relapse in schizophrenia and to develop model to predict relapse in each patient.

**Methods:** We have used classical Cox regression, survival penalized regression, as well as survival random forests to analyze clinical and inflammatory biological data collected in the network of the Schizophrenia Expert Centers in France in which individuals with SZ are clinically assessed and followed up annually for 3 years.

**Results:** Among 247 individuals with SZ, 71 (29 %) experienced a psychotic relapse during the 3-year follow-up period. The variables most consistently associated with relapses were smoking status, severity of positive symptoms and low global functioning. From a panel of inflammatory parameters, only IL-8 serum levels were associated with time to relapse. The predictive performance, assessed using C-index, was 0.54 using both penalized regression and random forests.

**Conclusions:** We found several clinical and biological variables consistently associated with relapses across three distinct statistical methods. However, despite these associations, the predictive capacity of these models remained low, highlighting that association does not necessarily mean prediction.

## 1. Introduction

Schizophrenia (SZ) is characterized by significant suffering, disability, and a substantial global economic burden. Despite advances in the understanding of its biological basis, the clinical management of psychotic disorders remains difficult. Up to 40 % of patients discharged from hospital are estimated to experience relapse within 1 year (Leucht et al., 2012; Emsley et al., 2013). The relapse rate within the first year after onset has been estimated to be approximately 35 %, and the lifetime risk of relapse up to 80 % (Porcelli et al., 2016). Relapse can have serious consequences such as worsening of symptoms, progressive cognitive deterioration, impaired functioning, loss of employment and strain on personal relationships (Almond et al., 2004). In addition, individuals experiencing a relapse frequently require hospitalization, which constitutes the principal source of SZ annual direct cost (Laidi et al., 2018).

To alleviate the burden of SZ on patients and their families, prolonging periods of effective symptom management is crucial. This goal could potentially be achieved through a personalized medicine approach, estimating the risk of relapse for each individual with SZ. As an initial step toward this objective, numerous authors have tried to identify clinical and biological factors associated with relapse in SZ individuals. To the best of our knowledge, specific biomarkers linked to relapse remain unidentified. Conversely, several clinical factors have been correlated with relapse, including male gender, younger age at onset, substance abuse, tobacco smoking, metabolic syndrome (Godin et al., 2018), low premorbid functioning, duration of untreated psychosis, shorter baseline hospitalization, poor adherence to treatment, limited insight into the illness and anger and aggressiveness (Alvarez-Jimenez et al., 2012; Hui et al., 2013; Porcelli et al., 2016; Fond et al., 2019). In contrast, the utilization of second-generation antipsychotics, as clozapine and olanzapine, as well as involvement in psychotherapy programs are associated with lower relapse rates (Efthimiou et al., 2024; Lähteenvuo et al., 2022; Tiihonen et al., 2019, 2017; Porcelli et al., 2016).

Immune abnormalities have repeatedly been found in subjects with schizophrenia. We and others have found that patients with schizophrenia have elevated blood cytokines, in particular when they are resistant to treatment (Leboyer et al., 2021). Genes involved in the regulation of the immune system, in particular those belonging to the human leukocyte antigen (HLA) cluster were demonstrated to modulate the risk of SZ. In particular, ancestral haplotypes inducing excessive synaptic pruning are associated with increased risk of schizophrenia (Tamouza et al., 2021; Tamouza et al., 2020). Among risk factors for

schizophrenia, prenatal maternal infection with a variety of different infectious agents and/or autoimmune disorders separately or in combination is a well-replicated factor (Brown, 2015; Benros et al., 2013). Significant improvement in patients with acute psychotic episode was reported in several studies when anti-inflammatory treatments were added to regular psychotropic (Fitton et al., 2022). Therefore, inflammatory factors may play a role in relapse that needs to be more specifically evaluated.

Despite the identification of several factors associated with relapse in SZ, algorithms that could be used by clinicians to predict relapse in an individual patient have not yet been discovered. This may be explained because most studies in the field did not go further than association. In a typical association study, a dataset is usually analyzed using univariate methods, such as parametric tests (correlations, *t*-test), or models such as linear or logistic regressions. A possible step further is to analyze the data in a survival framework. The time to an event data contains more clinical information than the simply binary dichotomy of event occurrence or absence, and it translates into more statistical information, as survival analyses tend to have greater statistical power than methods for binary outcomes such as logistic regression (George et al., 2014). While the Cox proportional hazards model is the most widely used, it is only suitable for small data sets and does not scale well to high dimensions (Stadtmauer et al., 2020). While regression methods allow for prediction, the model must be carefully validated to generate good estimates of the outcome using training and test or validation datasets (Iniesta et al., 2016). Training datasets are used to create the model while test or validation datasets must be used to confirm that the model is reliable. While the training and validation datasets could ideally come from independent populations, what is done in practice is to use resampling methods on the dataset to be derive a training, test, and validation datasets.

Based on the foregoing, in the present study, we aimed to produce a model to accurately predict relapse in individuals with SZ and to identify additional inflammatory markers associated with relapse. We focused on seven cytokines: interleukin (IL)-6, IL-7, IL-8, IL-10, IL-15, Tumor Necrosis Factor (TNF)- $\alpha$  and interferon (IFN)- $\gamma$ ; one Chemokine Ligand: (CCL)-2; and one neurotrophic factor: Brain Derived Neurotrophic Factor (BDNF). Their levels have been shown either to differ in schizophrenia patients when compared to healthy controls, associated with the severity of symptoms or with the clinical outcome. To name just a few of these studies, TNF-alpha levels before treatment were significantly higher in drug-naïve first episode of psychosis (FEP) patients than in chronic patients and healthy controls (Lin et al., 2021). Brain Derived Neurotrophic Factor (BDNF) has been linked to cognitive symptoms of

schizophrenia in many studies (Nieto et al., 2021). Patients with schizophrenia show raised serum levels of the pro-inflammatory chemokine CCL2 (Drexhage et al., 2008). In patients with First Episode Psychosis (FEP) exhibiting severe symptoms and treated with amisulpride, lower serum levels of interleukin (IL)-15 were associated with higher odds of being non-remitters 4 weeks after treatment (Martinuzzi et al., 2019).

To achieve this goal, we have used clinical and inflammatory biological data, collected in the network of the Schizophrenia Expert Centers in France, in which 247 patients with SZ were clinically assessed and followed-up annually (Schürhoff et al., 2015). We are extending previous work on prediction of relapse in this cohort (Fond et al., 2019) to the survival framework with the inclusion of additional biological variables.

## 2. Materials and methods

### 2.1. Study design and participants

The study sample was drawn from patients who were evaluated in hospitals belonging to the network of French Schizophrenia Expert Centers. The FACE-SZ cohort (FondaMental Academic Centers of Expertise for Schizophrenia) is based on a French national network of 10 Schizophrenia Expert Centers (Bordeaux, Clermont-Ferrand, Colombes, Créteil, Grenoble, Lyon, Marseille, Montpellier, Strasbourg, and Versailles). All clinically stable outpatients, defined by no hospitalizations or changes in treatment during the 4 weeks before evaluation, aged 16 years and older with a DSM-IV-TR diagnosis of schizophrenia or schizoaffective disorder were included in this study, which was conducted from 2014 until the end of 2019. As this sample aimed to be representative of a stabilized community-dwelling national sample of outpatients with schizophrenia, unremitted patients with moderate to severe symptoms were also included in the Schizophrenia Expert Centers network. All patients were referred by their general practitioner or psychiatrist, who subsequently received a detailed evaluation report with suggestions for personalized interventions. Patients were evaluated with a thorough, standardized assessment and were followed up every year for 3 years. Full description of population, characteristics and treatments can be found elsewhere (Schürhoff et al., 2019). A non-opposition form was signed by all participants, as this study included data coming from regular care assessments. All data were collected anonymously. The assessment protocol was approved by the relevant ethical review board (CPP-Ile de France IX, January 18, 2010). Access to the system is carefully regulated and approval was obtained from the ethical committee and the committee in charge of the safety of computerized databases (CNIL; DR-2012-157).

All patients were interviewed by members of the specialized multidisciplinary team of the Expert Center. They were interviewed by a trained psychiatrist who used the Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Patient Edition (SCID-I/P), to confirm the diagnosis. Alcohol consumption, tobacco smoking, current cannabis use, substance abuse, eating disorder and current psychotropic treatments were recorded. Information concerning the onset and course of the illness were collected. Schizophrenia symptomatology was assessed using the Positive and Negative Syndrome Scale (PANSS). Current depressive symptoms were evaluated using the Calgary Depression Scale for Schizophrenia (CDSS). Medication adherence was measured using the Medication Adherence Rating Scale (MARS) (Fond et al., 2017). The sum of the 10 items yields a final score ranking from 0 (poorest adherence to treatment) to 10 (best adherence to treatment). The first four items were related to “medication adherence behavior”, the following four items represented the “subjects’ attitudes toward taking medication.” The remaining two items represented “subjective negative side effects”. Exposure to maltreatment during childhood was measured using the Childhood Trauma Questionnaire (CTQ). Psychological, social, and occupational functioning was measured using the Global Assessment of Functioning (GAF) scale. Body mass index (BMI) was calculated

as the weight in kilograms divided by the square of the height in meters. The presence of metabolic syndrome (MetS) was defined according to the criteria of the International Diabetes Federation, which requires that 3 or more of the following 5 criteria be met: high waist circumference (> 94 cm for men and > 80 cm for women), hypertriglyceridemia ( $\geq 1.7$  mmol/L or on lipid-lowering medication), low HDL cholesterol level (< 1.03 mmol/L in men and < 1.29 mmol/L in women), high blood pressure ( $\geq 130/85$  mmHg or on antihypertensive medication), and high fasting glucose concentration ( $\geq 5.6$  mmol/L or on glucose-lowering medication).

Psychotic relapse was defined by a psychotic acute episode of at least 7 consecutive days in the last 12 months (Siafis et al., 2024). To avoid memory bias, relatives and treating psychiatrists were also interviewed to report any potential psychotic relapse during the follow-up that may not have been reported by the patient. Other definitions of relapse such as a hospitalization only or a combination of hospitalization and psychotic acute episode have also been tested.

### 2.2. Blood biomarkers

To measure blood analytes, venous blood was obtained from fasting subjects between 7:00 am and 9:00 am. Routine blood samples were used to measure levels of triglycerides and glucose (requiring a 10-h fast), as well as low-density lipoprotein, high-density lipoprotein (HDL) cholesterol and total cholesterol. Serum levels of interleukin (IL)-6, IL-7, IL-8, IL-10, IL-15, interferon (IFN)-gamma, Tumor Necrosis Factor (TNF)- $\alpha$ , Chemokine Ligand (CCL)-2, and Brain Derived Neurotrophic Factor (BDNF) were measured using a MesoScale Discovery (MSD) human V-Plex electro-chemiluminescence assay (MSD, Rockville, Maryland, USA, and data were already reported in previous studies (Leboyer et al., 2021; Foiselle et al., 2022).

### 2.3. Data analysis

Categorical variables were summarized using percentages and quantitative variables using means (standard deviation). Sociodemographic and clinical characteristics, as well as inflammatory parameters of individuals with or without psychotic relapse during follow-up were compared using chi-square tests for categorical variables and Student's or Mann-Whitney tests (depending on the distribution of the variables) for continuous variables.

We subsequently used survival framework methodologies with the aim of predicting relapse while also evaluating the importance of inflammatory parameters. We used the same panel of variables in all analysis. We used three different approaches: classical Cox regression, penalized Cox regression and survival random forests.

The penalized regression framework uses linear models with penalties to avoid extreme parameters that may cause overfitting, while simultaneously performing variable selection and addressing the issue of multicollinearity. We have chosen the Elastic net regression which is well suited when the number of variables is larger than the number of individuals and will select in tandem any group of correlated variables (Stadtmauer et al., 2020). For each variable, we computed the variable inclusion probability (VIP, % runs in which the variable was selected by the penalized regression), the median hazard ratios and the percentile bootstrap confidence intervals (CI). Missing data were imputed using the Multivariate Imputation (MI) by Chained Equations (MICE) procedure in R (van Buuren and Groothuis-Oudshoorn, 2011) and non-parametric bootstrap was used for statistical inference (Abram et al., 2016). The CARET (Kuhn, 2008), GLMNET (Simon et al., 2011) and Survival R packages were used to implement Elastic Net penalized Cox regression models and the classical Cox regression.

We also applied the random survival forests nonparametric ensemble method, that is an extension to censored survival data of the random forest technique. It gathers the results from a great number of decision trees, generated from resampling the data, recursively partitioning the

covariate space using binary splits to form groups of subjects who are similar according to the survival outcome (Pickett et al., 2021). The method can handle multiple covariates as well as complex, nonlinear relationships between covariates without need for prior specification. Missing data was imputed as part of the random forest algorithm itself

(Wright and Ziegler, 2017). The different models' performance was assessed using Harrell's C-index, a metric very commonly used for the global evaluation of prognostic models in survival analysis. It summarizes how good a predicted risk score describes an observed sequence of events and ranges from 0 to 1. There are not clearly defined thresholds

**Table 1**

Associations between clinical and inflammatory biological variables and relapse determined using univariate and multivariable Cox regressions. The number and frequency of missing data are shown for individual variable. Socio-demographic, clinical and inflammatory biological patient characteristics are shown for all patients, for the patients who have relapsed and for those who have not. Hazard Ratios and *p*-values are indicated when the dataset was analyzed using univariate and multivariable Cox regression. PPANSS, Positive and Negative Syndrome Scale (PANSS) positive subscale; NPANSS, PANSS negative subscale; GPANSS, General Psychopathology subscale of the PANSS; MI, multiple imputation.

	Descriptive				Cox regression		Cox regression		
	Missing Nb (%)	All (N = 247)	Relapse		Univariate analysis MI		Multivariable analysis MI		
			No (n = 176)	Yes (n = 71)	HR	<i>p</i> -value	HR	<i>p</i> -value	
	Mean (SD) or Nb (%)		71.30 %	28.70 %					
Survival time		15.6 (9.2)							
Age (Years)	0 (0.0)	29.8 (8.2)	29.31 (8.03)	31.24 (8.38)	1.027	0.05	1.006	0.748	
Body mass index (kg/m <sup>2</sup> )	9 (3.6)	26.8 (5.3)	26.61 (5.45)	27.44 (4.98)	1.023	0.293			
Gender									
Male	0 (0.0)	194 (78.2)	139 (72.02 %)	54 (27.98 %)	0.81	0.452			
Female		54 (21.8)	37 (68.52 %)	17 (31.48 %)	1 (ref)				
Education (Years)	27 (10.9)	12.7 (2.4)	12.8 (2.37)	12.3 (2.54)	0.924	0.142			
Metabolic syndrome									
No	33 (13.3)	159 (74.0)	113 (71.52 %)	45 (28.48 %)	1 (ref)				
Yes		56 (26.0)	39 (69.64 %)	17 (30.36 %)	1.127	0.658			
Substance abuse									
No	0 (0.0)	121 (48.8)	92 (76.03 %)	29 (23.97 %)	1(ref)				
Yes		127 (51.2)	84 (66.67 %)	42 (33.33 %)	1.29	0.296			
Alcohol									
No	7 (2.8)	161 (66.8)	118 (73.29 %)	43 (26.71 %)	1(ref)				
Yes		80 (33.2)	52 (65.00 %)	28 (35.00 %)	1.258	0.353			
Cannabis									
No	3 (1.2)	153 (62.4)	110 (71.90 %)	43 (28.10 %)	1(ref)				
Yes		92 (37.6)	65 (70.65 %)	27 (29.35 %)	0.911	0.705			
Tobacco									
No	6 (2.4)	98 (40.5)	76 (77.55 %)	22 (22.45 %)	1(ref)		1 (ref)		
Past		19 (7.9)	9 (47.37 %)	10 (52.63 %)	2.78	0.009	2.646	0.02	
Current		125 (51.7)	86 (69.35 %)	38 (30.65 %)	1.427	0.178	1.248	0.435	
Suicide attempts									
No	5 (2.0)	198 (81.5)	144 (73.10 %)	53 (26.90 %)	1(ref)				
Yes		45 (18.5)	28 (62.22 %)	17 (37.78 %)	1.44	0.195			
Age first psychotic episode (y)	20 (8.1)	20.5 (5.2)	20.12 (4.47)	21.69 (6.53)	1.04	0.078	1.043	0.15	
Nb psychotic episodes	0 (0.0)	10.8 (9.2)	10.69 (9.33)	11.11 (8.99)	1.009	0.497			
Illness duration (Months)	20 (8.1)	9.3 (7.1)	8.9 (7.13)	10.22 (6.87)	1.016	0.326			
Childhood trauma	11 (4.4)	40.9 (11.3)	39.99 (10.32)	43.21 (13.31)	1.021	0.037	1.009	0.366	
Medication adherence score	14 (5.6)	6.4 (2.2)	6.41 (2.13)	6.29 (2.3)	0.97	0.59			
Lifetime hospitalization									
No	7 (2.8)	52 (21.6)	39 (75.00 %)	13 (25.00 %)	1 (ref)				
Yes		189 (78.4)	132 (70.21 %)	56 (29.79 %)	1.272	0.437			
PPANSS	10 (4.0)	14.2 (5.5)	13.35 (5.11)	16.27 (6.03)	1.066	0.002	1.03	0.31	
NPANSS	9 (3.6)	17.6 (7)	17.12 (6.82)	18.63 (7.3)	1.016	0.325			
GPANSS	9 (3.6)	32.9 (10)	31.58 (9.11)	36.23 (11.26)	1.034	0.005	1.008	0.649	
Functioning	9 (3.6)	52 (15.3)	54.39 (14.78)	46.34 (15.31)	0.974	0.003	0.987	0.287	
Depressive symptoms	6 (2.4)	3.7 (3.9)	3.37 (3.54)	4.54 (4.7)	1.055	0.064	1.028	0.375	
Antidepressants									
No	18 (7.3)	165 (71.7)	121 (73.78 %)	43 (26.22 %)	1 (ref)				
Yes		65 (28.3)	43 (66.15 %)	22 (33.85 %)	1.228	0.431			
Anticonvulsants									
No	18 (7.3)	200 (87.0)	146 (73.37 %)	53 (26.63 %)	1 (ref)				
Yes		30 (13.0)	18 (60.00 %)	12 (40.00 %)	1.771	0.076	1.652	0.135	
Benzodiazepines									
No	16 (6.5)	188 (81.0)	138 (73.80 %)	49 (26.20 %)	1 (Ref)				
Yes		44 (19.0)	27 (61.36 %)	17 (38.64 %)	1.39	0.235			
IL-6 (pg/ml)	1 (0.4)	2.2 (9.5)	2.22 (9.8)	2.18 (9.25)	0.999	0.951			
IL-7 (pg/ml)	0 (0.0)	13.2 (6.3)	13.6 (6.36)	12.36 (6.19)	0.975	0.213			
IL-8 (pg/ml)	7 (2.8)	45.6 (143.6)	52.77 (164.96)	24.03 (46.91)	0.998	0.312			
IL-10 (pg/ml)	1 (0.4)	0.5 (0.7)	0.49 (0.52)	0.49 (0.93)	0.994	0.975			
IL-15 (pg/ml)	1 (0.4)	2.5 (0.7)	2.48 (0.79)	2.41 (0.5)	0.925	0.64			
TNF-α (pg/ml)	1 (0.4)	4.9 (12.4)	4.84 (12.07)	5.18 (13.18)	1.001	0.894			
IFN-γ (pg/ml)	1 (0.4)	7.4 (40.3)	8.52 (47.71)	4.43 (4.69)	0.984	0.481			
BDNF (pg/ml)	1 (0.4)	15,154.1 (15,296.2)	14,990.33 (15,037.87)	15,560.04 (16,021.06)	1	0.801			
CCL2 (pg/ml)	2 (0.8)	294.6 (110.9)	288.83 (114.12)	304.91 (107.25)	1.001	0.528			

for goodness of fit evaluation, however, a common interpretation is that a value of 0.5 means that the model is no better than random chance, over 0.7 indicates a good model and over 0.8 indicates a strong model. It has several pitfalls, primarily that it remains time dependent, so it is not necessarily the best metric (Longato et al., 2020), and Harrell's estimator not the only implementation, but it is the metric implemented in R in common for all the methods we used.

### 3. Results

Our study sample consisted of 247 patients with schizophrenia, among whom 71 (28.7 %) experienced at least one psychotic relapse during the follow-up interval and 176 (71.3 %) did not. The mean duration of follow-up was 20 months and mean time to relapse was 15.6 months (Supplementary Fig. 1). When first comparing the patients from these two groups (Table 1, significance threshold  $p$ -value  $<0.1$ ) we found that relapsing individuals were older, were more frequently smokers, exhibited higher positive and general symptomatology symptoms, lower general functioning, and higher levels of IL-10 in serum (Table 1, significance threshold  $p$ -value  $<0.1$ ).

Subsequently, while using the univariate Cox regression, we identified age, tobacco, age of first psychotic episode, positive and general symptomatology, global assessment of functioning, depressive symptoms and use of anticonvulsants associated with time to relapse. None of the tested inflammatory parameters was associated with time to relapse. Only tobacco remained significantly associated with time to relapse in the multivariable Cox regression model.

We then analyzed the same dataset using Cox penalized regression (Table 2). Considering a conservative VIP threshold of 80 %, we observed that sex, education, alcohol, cannabis, history of tobacco smoking, age of first psychotic episode, childhood trauma, positive and general symptomatology, general functioning, anticonvulsants, and serum levels of IL-8 were associated with time to relapse (Table 2). The predictive capability of the models, assessed using the C-index, was 0.54.

Finally, we analyzed the dataset using survival random forests to identify the most important variables at predicting relapse (Fig. 1). This figure depicted the variable importance determined by shuffling its values and observing the impact on model accuracy. In decreasing order of importance were global assessment of functioning, tobacco, positive and depressive symptoms, age of first psychotic episode, and IL-8 serum levels were found in a decreasing order of importance. The predictive capability of the model assessed using the C-index was also 0.54.

Taken together, we observed that variables associated with relapse differed according to the statistical method and significance threshold. Among the 24 tested variables, being a former smoker seems to be the one most consistently associated with relapse ( $p$ -value  $<0.05$  using both univariate and multivariable Cox regression; VIP  $> 90$  % using penalized regression: second most important variable using random forests). General functioning, and positive symptoms were also consistently associated with relapse using univariate Cox regression ( $p$ -value  $<0.1$ ) and penalized regression (VIP  $> 85$  %) and were the first and the fourth most important variables using random forests respectively. General symptomatology was associated with relapse using both univariate Cox regression ( $p < 0.1$ ) and penalized regression (VIP  $> 80$  %) and was the seventh most important variable using random forests. Age of first psychotic episode was associated with relapse using univariate Cox regression ( $p < 0.1$ ), and penalized regression (VIP  $> 85$  %) and was the sixth most important variable using random forests. Depression symptoms were associated with relapse using univariate Cox regression ( $p < 0.1$ ), and penalized Cox regression (VIP  $> 75$  %) and was the fourth most important variable using random forests. Among the other clinical variables, anticonvulsants, age, sex, education, and cannabis were also associated with relapse but not consistently according to the different statistical methods. As for the nine inflammatory biological variables that were tested, IL-8 was associated with relapse using penalized

**Table 2**

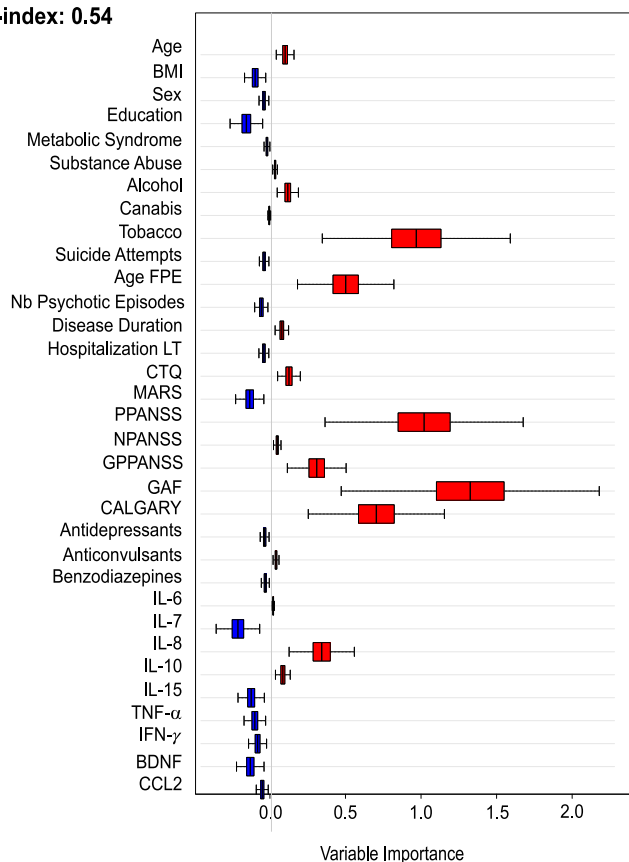
Association between clinical and inflammatory biological variables and relapse measured using penalized Cox regression. Data were obtained on 40 multiple imputed (MI) datasets and for 1000 Bootstraps in each imputation. Hazard Ratios, 95 % Confidence Interval (CI) and Variable Inclusion Probability (VIP) are shown. Variables with a VIP  $>0.80$  are highlighted. PPANSS, Positive and Negative Syndrome Scale (PANSS) positive subscale; NPANSS, PANSS negative subscale; GPANSS, General Psychopathology subscale of the PANSS.

	Hazard ratio	95 %CI	VIP
Age	1.012	[0.483,1.848]	0.695
Body mass index (BMI)	1.014	[0.933,1.1]	0.745
Sex Female	1.362	[0.867,3.494]	0.806
Sex Male	0.832	[0.511,1.066]	0.801
Education	0.932	[0.776,1.086]	0.818
Metabolic syndrome No	1.142	[0.635,2.39]	0.739
Metabolic syndrome Yes	0.937	[0.606,1.257]	0.733
Substance abuse No	0.706	[0.19,1.405]	0.754
Substance abuse Yes	1.231	[0.844,2.571]	0.75
Alcohol No	0.802	[0.32,1.545]	0.796
Alcohol Yes	1.14	[0.813,1.857]	0.789
Canabis No	1.667	[0.981,5.766]	0.83
Canabis Yes	0.727	[0.352,1.004]	0.825
Tobacco Past	2.981	[1.111,12.009]	0.909
Tobacco Yes	0.958	[0.399,1.96]	0.337
Tobacco No	0.684	[0.272,1.041]	0.634
Suicide attempts No	0.948	[0.485,2.163]	0.765
Suicide attempts Yes	1.022	[0.686,1.548]	0.762
Age First psychotic episode (FPE)	1.051	[0.626,2.087]	0.871
Nb psychotic episodes	0.997	[0.939,1.034]	0.743
Disease duration	0.999	[0.501,2.3]	0.552
Lifetime hospitalization No	0.934	[0.442,1.822]	0.744
Lifetime hospitalization Yes	1.029	[0.725,1.587]	0.735
Childhood Trauma Questionnaire (CTQ)	1.013	[0.986,1.047]	0.818
Medication adherence score (MARS)	1.053	[0.847,1.324]	0.746
PPANSS	1.023	[0.947,1.104]	0.883
NPANSS	0.956	[0.881,1.01]	0.743
GPPANSS	1.025	[0.983,1.103]	0.838
Global Assessment of Functioning (GAF)	0.976	[0.929,1]	0.912
Depressive symptoms (CALGARY)	1.022	[0.903,1.137]	0.779
Antidepressants No	1.025	[0.61,1.982]	0.728
Antidepressants Yes	0.999	[0.699,1.359]	0.723
Anticonvulsant No	0.805	[0.37,1.586]	0.828
Anticonvulsant Yes	1.141	[0.81,1.796]	0.815
Benzodiazepines No	1.109	[0.638,2.328]	0.744
Benzodiazepines Yes	0.961	[0.638,1.28]	0.742
IL-6	1.011	[0.656,1.191]	0.669
IL-7	0.982	[0.915,1.044]	0.78
IL-8	0.998	[0.989,1.001]	0.829
IL-10	1.136	[0.284,2.654]	0.772
IL-15	0.886	[0.502,1.811]	0.771
TNF- $\alpha$	1.005	[0.875,1.142]	0.641
IFN- $\gamma$	0.996	[0.911,1.04]	0.779
BDNF	1	[1,1]	0.73
CCL2	1	[0.997,1.004]	0.727

regression (VIP  $> 80$  %) and was the sixth most important variable using random forests.

When two other definitions of relapse were tested (hospitalization alone or the combination of h), we found that the results for clinical factors were generally consistent with those from the first definition, although some factors, such as a history of suicide attempts, emerged as predictors. Regarding inflammatory markers, BDNF was identified as important when relapse was defined solely by hospitalization, whereas TNF- $\alpha$  and IL-7 were relevant for the combined definition of relapse (hospitalization + acute episode). However, similar to IL-8, their estimates were close to 1, reinforcing the idea that inflammatory markers are likely not reliable biomarkers for predicting the severity of schizophrenia (results not shown).

C-index: 0.54



**Fig. 1.** Variable importance determined using survival random forests. The variable importance of individual clinical and inflammatory biological variables is shown. The importance of a variable is determined by shuffling its values and observing the impact on model accuracy. The presence of values to the right of zero (with red colors) means that these variables increase model accuracy when included, contribute positively to prediction performance, higher values indicating more important predictors. On the opposite, presence of values to the left of zero (with blue colors) means that variables decrease model accuracy when included and suggests that removing the variable might improve the model. BMI, Body mass Index; FPE, First Psychotic Episode; Hospitalization LT, Lifetime Hospitalization; CTQ, Childhood Trauma Questionnaire; MARS, Medication Adherence Rating Scale; PPANSS, Positive and Negative Syndrome Scale (PANSS) positive subscale; NPANSS, PANSS negative subscale; GPPANSS, General Psychopathology subscale of the PANSS; GAF, Global Assessment of Functioning scale; CALGARY, Calgary Depression Scale for Schizophrenia; (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

#### 4. Discussion

The main objective of this study was to develop a model to predict relapse and to identify clinical and inflammatory parameters associated with relapse in schizophrenia. To this aim, we have used several methods: classical univariate and multivariable Cox regressions, multivariable Cox penalized regression and survival random forests to analyze clinical and inflammatory biological data, collected from 247 patients with schizophrenia, which have been followed annually for 3 years in the FondaMental Expert Centers.

Our results showing that substance use disorders, particularly alcohol and cannabis, as well as lower functioning and higher depressive symptoms, were risk factors for relapse confirm the extent previous literature on clinical risk for relapse in schizophrenia (Porcelli et al., 2016; Lecomte et al., 2019; Abdel-Baki et al., 2011). Further, data are also consistent about the identification of lower education level as a risk factor for poor outcome. Consistently, several studies have shown that

pre-morbid function, including the educational attainment, the work position and the social adjustment before the onset of SCZ symptoms, are strongly associated with relapse (Dunn et al., 2008; Maramis et al., 2022; Fond et al., 2019). Male gender was found to be associated with a higher risk of relapse on most studies (Grossman et al., 2008; Porcelli et al., 2016), which is not consistent with our results. While the association between smoking and schizophrenia is known for several decades (Ding and Hu, 2021), Schoretsanitis et al. and our study are the firsts to report that it is an important risk factor for time to relapse too (Schoretsanitis et al., 2022). However, our study failed to identify adherence to medication as a factor associated with time to relapse whereas poor adherence to oral antipsychotic medications has been largely demonstrated in the literature.

In the present study, only IL-8 was found to be a predictor of time to relapse. Such observation could be explained by the dual function of IL-8 which is both a cytokine and a chemoattractant known to attract immune effector cells toward a site of inflammation. The difference in IL-8 levels between relapsers and non-relapsers could be interpreted in two ways. First, IL-8 could serve as a state biomarker associated with the probability of relapse. Alternatively, this association may suggest that changes in IL-8 could indicate that relapse is on the way. Only two recent studies, to our knowledge, have investigated whether inflammatory parameters could predict relapse in longitudinal analysis including SZ individuals and found no significant associations (Pillai et al., 2018; Miller et al., 2023). In a longitudinal cohort of 221 individuals with schizophrenia, followed-up for 30 months, Pillai et al. found no association between BDNF, the only markers studied, and the risk of relapse. More recently, using the same cohort, Miller et al. found that IL-2, IL-4, IL-6, IL-8, IL-10, tumor necrosis factor (TNF)- $\alpha$ , IFN- $\gamma$ , and granulocyte-macrophage colony-stimulating factor (GM-CSF) were not predictive markers of relapse. While we observed a statistically significant association between changes in IL-8 levels and relapse, it should be noted that the effect size was extremely low and may not reflect any meaningful biological or clinical difference. Even if it did, it may not be clinically useful.

As for the predictive capabilities of the different models, the C-index were 0.54 for penalized Cox regression and random forests model. While the statistical methods that we have used allowed for identifying variables consistently associated with relapse in schizophrenia, the predictive capabilities of these models were low. Previous work on this dataset also achieved modest prediction metrics (accuracy 63.8 %, sensitivity 71.0 % and specificity 44.8 %) with a different technique and panel of predictors (Fond et al., 2019). It shows that associations do not by themselves support conclusions outside of the recruitment setting. In many cases, unusually strong associations only offer the allure of associations that may generalize to new individuals and contexts. Prediction raises the bar in aiming at successful generalization, by building a model to encapsulate predictive principles from one cohort (or train dataset) and then explicitly testing for generalization by model-derived quantities in every individual of a second, unseen cohort (or test dataset). This goal of clinical outcome prediction in patients who have not yet been observed is not well aligned with the focus of replication analysis.

Conflating association and prediction have generated much misunderstanding and myriads of studies irrelevant for clinical decision-making (Bzdok and Ioannidis, 2019). Strategies to evaluate trustworthy generalizations beyond the model development setting are strategies of criticizing whether associations can predict clinical end points in new individuals. This study reinforces the statement that “association is not prediction” as emphasized by several authors in the field on predictive medicine (Bzdok et al., 2021).

Our study has several strengths including a longitudinal design, the use of both clinical and inflammatory parameters and used a diverse range of prediction methodologies. Despite the effort in collecting the data, the study dataset consisted of 247 individuals among whom 71 relapsed. There is a clear imbalance in the outcome classes, with 28.7 % of patients relapsing and 71.3 % not. The problem with unbalanced

datasets is that most evaluation metrics, including the C-Index, may be unable to portrait the real performance of the model. The model will display a tendency to classify every individual in the majority class, leading to a high-performance metrics (Shi et al., 2022). There are methodologies such as the Synthetic Minority Oversampling Techniques Nominal Continuous (SMOTE-NC) (Shi et al., 2022) and the bootstrap-based technique Random Over-Sampling Examples (ROSE) (Menardi and Torelli, 2014) that artificially generate new samples of the minority classes. We implemented both methodologies (results not shown) without gain or loss in terms of predictive performance.

Additionally, the number of predictors variables available in the present dataset is substantial, in particular if we consider the number of individuals and the number of events, with a ratio of 3 events per variable. Classic theory of statistical models is well established if the set of independent variables is small, however, if we have a large number of candidate variables (Chowdhury and Turin, 2020), the results of classical statistics analysis need to be interpreted with caution (Vittinghoff and McCulloch, 2007). We needed therefore to establish a variable selection strategy. The classical method in epidemiology is to start with a univariate analysis, implementing a regression model of choice for each variable, and using a pre-defined significance threshold to select variables to proceed for a multivariable step. There is an extended literature discussing the dangers of using any of the forward, backwards or step-wise selection methodologies (Smith, 2018) and therefore we refrained from doing so. Relatively recent theoretical approaches such as penalized regression methods, perform variable selection while handling important issues of multicollinearity and selection bias, that mask the capabilities of variables to be identified as important, and are still relatively straightforward to interpret (Greenwood et al., 2020). We implemented the classical univariate strategy, a penalized version of Cox regression and Random forests (a prediction orientated framework), to highlight disparities in the results. Another limitation of this work is that we did not consider other variables that could be potentially associated with time to relapse, such as those related to cognition, the potential role of air pollution (Pignon et al., 2022) and other environmental variables, neither did we explore the existence of competing risks to relapse, the main event.

#### 4.1. Conclusion

To summarize, we have found numerous clinical and biological variables which were consistently associated with relapse across three distinct statistical methods. However, despite these associations, the predictive capacity of these models remained low, highlighting a significant disparity between association and prediction.

#### Ethical standards

The study was carried following ethical principles for medical research involving humans (WMA, Declaration of Helsinki) and the French Jarde law. The assessment protocol was approved by the relevant ethical review board (CPP-Ile de France IX, January 18, 2010). All data were collected anonymously. As this study includes data coming from regular care assessments, a non-opposition form was signed by all participants.

#### CRediT authorship contribution statement

**Susana Barbosa:** Writing – original draft, Methodology, Formal analysis. **Ryad Tamouza:** Writing – original draft, Supervision, Funding acquisition, Conceptualization. **Marion Leboyer:** Writing – original draft, Validation, Supervision, Resources, Funding acquisition, Conceptualization. **Bruno Aouizerate:** Writing – original draft, Validation, Resources, Data curation. **Christelle Andrieu:** Validation, Resources, Data curation, Conceptualization. **Myrtille Andre:** Validation, Resources, Data curation, Conceptualization. **Wahid Boukouaci:**

Validation, Resources, Methodology, Data curation, Conceptualization. **Delphine Capdevielle:** Validation, Resources, Data curation, Conceptualization. **Isabelle Chereau:** Validation, Resources, Data curation, Conceptualization. **Julie Clauss Kobayashi:** Validation, Resources, Data curation, Conceptualization. **Nathalie Coulon:** Validation, Resources, Data curation, Conceptualization. **Jean-Michel Dorey:** Validation, Resources, Data curation, Conceptualization. **Laetitia Davidovic:** Validation, Methodology, Conceptualization. **Caroline Dubertret:** Validation, Resources, Data curation, Conceptualization. **Eric Fakra:** Validation, Resources, Data curation, Conceptualization. **Guillaume Fond:** Validation, Resources, Data curation, Conceptualization. **Tudi Goze:** Validation, Resources, Data curation, Conceptualization. **Olfa Khalfallah:** Validation, Resources, Data curation, Conceptualization. **Sylvain Leignier:** Validation, Resources, Data curation, Conceptualization. **Pierre Michel Llorca:** Validation, Resources, Data curation, Conceptualization. **Jasmina Mallet:** Validation, Resources, Data curation, Conceptualization. **Emanuela Martinuzzi:** Validation, Resources, Data curation, Conceptualization. **David Misdrahi:** Validation, Resources, Data curation, Conceptualization. **Nicolas Oriol:** Validation, Resources, Data curation, Conceptualization. **Baptiste Pignon:** Writing – original draft, Validation, Resources, Data curation, Conceptualization. **Romain Rey:** Validation, Resources, Data curation, Conceptualization. **Paul Roux:** Validation, Resources, Data curation, Conceptualization. **Franck Schürhoff:** Validation, Resources, Data curation, Conceptualization. **Benoit Schorr:** Validation, Resources, Data curation, Conceptualization. **Mathieu Urbach:** Validation, Resources, Data curation, Conceptualization. **Etienne Very:** Validation, Resources, Data curation, Conceptualization. **Ching-Lien Wu:** Validation, Resources, Data curation, Conceptualization. **Michael Benros:** Validation, Funding acquisition. **Judit Simon:** Validation, Funding acquisition. **Alkomiet Hasan:** Validation, Funding acquisition. **Nicolas Glaichenhaus:** Writing – original draft, Validation, Supervision, Methodology, Funding acquisition, Conceptualization. **Ophélie Godin:** Writing – original draft, Validation, Supervision, Methodology, Conceptualization.

#### Declaration of competing interest

The authors declare that they have no competing interests.

#### Acknowledgments

We thank all the patients who participated in the study. The work was funded by grants from the International IF foundation (RT), Fondation de France (NG), Fondation FondaMental and the Université Côte d'Azur (SB). This project was supported by the Agence National de la Recherche (ANR) under the frame of ERA PerMed and by the ANR under reference ANR-11-IDEX-0004-02 and ANR-10-COHO-10-01. Support from the LABEX BioPSY and the LABEX SIGNALIFE (ANR-11-LABX-0028-01). This study is supported by the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF, FKZ: 01KU2002), the Innovation Fund Denmark, the Austrian Science Fund (Wissenschaftsfonds, FWF), and the ANR under the project title "PerMedSchiz: Personalized medical treatment for schizophrenia. \*List of FondaMental Advanced Center of Expertise (FACE-SZ) collaborators: Andre M<sup>4</sup>, Andrieu-Haller C<sup>5</sup>, Aouizerate B<sup>6,7</sup>, Berna F<sup>10</sup>, Blanc O<sup>9</sup>, Capdevielle D<sup>4</sup>, Cléry M<sup>1,2,3</sup>, Chereau-Boudet I<sup>9</sup>, Clauss-Kobayashi, J<sup>10</sup>, Coulon N<sup>11</sup>, Dassing R<sup>10</sup>, Dorey JM<sup>12</sup>, Dubertret C<sup>13,14</sup>, D'Amato T<sup>12</sup>, Esselin A<sup>18</sup>, Fakra E<sup>15</sup>, Fond G<sup>5</sup>, Gabayet F, Goze T<sup>16</sup>, Jarroir M<sup>18</sup>, Lacelle D<sup>9</sup>, Leboyer M<sup>1,2,3</sup>, Leignier S<sup>11</sup>, Llorca PM<sup>9</sup>, Mallet J<sup>13,14</sup>, Metairie E<sup>5</sup>, Michel T<sup>4</sup>, Misdrahi D<sup>17</sup>, Oriol N<sup>15</sup>, Passerieux C<sup>18</sup>, Pignon B<sup>1,2,3</sup>, Peri P, Portulier C<sup>13,14</sup>, Rey R<sup>12</sup>, Roman C<sup>11</sup>, Roux P<sup>18</sup>, Schorr B<sup>10</sup>, Schürhoff F<sup>1,2,3</sup>, Szöke A<sup>1,2,3</sup>, Tessier A<sup>6</sup>, Urbach M<sup>18</sup>, Very E<sup>16</sup>, Zinetti-Bertschy A<sup>10</sup>

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.pnpbp.2025.111304>.

## Data availability

Data will be made available on request.

## References

- Abdel-Baki, A., Lesage, A., Nicole, L., Cossette, M., Salvat, E., Lalonde, P., 2011. Schizophrenia, an illness with bad outcome: myth or reality? *Can. J. Psychiatry Rev. Can. Psychiatr.* 56 (2), 92–101.
- Abram, S.V., Helwig, N.E., Moodie, C.A., DeYoung, C.G., MacDonald, A.W., Waller, N.G., et al., 2016. Bootstrap enhanced penalized regression for variable selection with neuroimaging data. *Front. Neurosci.* 10, 344.
- Almond, S., Knapp, M., Francois, C., Toumi, M., Brugha, T., 2004. Relapse in schizophrenia: costs, clinical outcomes and quality of life. *Br. J. Psychiatry* 184 (4), 346–351.
- Alvarez-Jimenez, M., Priede, A., Hetrick, S.E., Bendall, S., Killackey, E., Parker, A.G., et al., 2012. Risk factors for relapse following treatment for first episode psychosis: a systematic review and meta-analysis of longitudinal studies. *Schizophr. Res.* 139 (1–3), 116–128.
- Benros, M.E., Waltoft, B.L., Nordentoft, M., Ostergaard, S.D., Eaton, W.W., Krogh, J., et al., 2013. Autoimmune diseases and severe infections as risk factors for mood disorders: a nationwide study. *JAMA Psychiatry* 70 (8), 812–820.
- Brown, A.S., 2015. The Kraepelinian dichotomy from the perspective of prenatal infectious and immunologic insults. *Schizophr. Bull.* 41 (4), 786–791.
- Bzdok, D., Ioannidis, J.P.A., 2019. Exploration, inference, and prediction in neuroscience and biomedicine. *Trends Neurosci.* 42 (4), 251–262.
- Bzdok, D., Varoquaux, G., Steyerberg, E.W., 2021. Prediction, not association, paves the road to precision medicine. *JAMA Psychiatry* 78 (2), 127–128.
- Chowdhury, M.Z.I., Turin, T.C., 2020. Variable selection strategies and its importance in clinical prediction modelling. *Fam. Med. Commun. Health* 8 (1), 262.
- Ding, J.B., Hu, K., 2021. Cigarette smoking and schizophrenia: etiology, clinical, pharmacological, and treatment implications. *Schizophr. Res. Treat* 2021.
- Drexhage, R.C., Padmos, R.C., de Wit, H., Versnel, M.A., Hooijkaas, H., der Lely, A.J. van, et al., 2008. Patients with schizophrenia show raised serum levels of the pro-inflammatory chemokine CCL2: association with the metabolic syndrome in patients? *Schizophr. Res.* 102 (1–3), 352–355.
- Dunn, E.C., Wewiorski, N.J., Rogers, E.S., 2008. The meaning and importance of employment to people in recovery from serious mental illness: results of a qualitative study. *Psychiatr. Rehabil. J.* 32 (1), 59–62.
- Efthimiou, O., Taipale, H., Radua, J., Schneider-Thoma, J., Pinzón-Espinosa, J., Ortuño, M., et al., 2024. Efficacy and effectiveness of antipsychotics in schizophrenia: network meta-analyses combining evidence from randomised controlled trials and real-world data. *Lancet Psychiatry* 11 (2), 102–111.
- Emsley, R., Chiliza, B., Asmal, L., Harvey, B.H., 2013. The nature of relapse in schizophrenia. *BMC Psychiatry* 13, 50.
- Fitton, R., Sweetman, J., Heseltine-Carp, W., van der Feltz-Cornelis, C., 2022. Anti-inflammatory medications for the treatment of mental disorders: a scoping review. *Brain Behav. Immun.* Health 26.
- Foiselle, M., Barbosa, S., Godin, O., Wu, C.L., Boukouaci, W., Andre, M., et al., 2022. Immuno-metabolic profile of patients with psychotic disorders and metabolic syndrome. Results from the FACE-SZ cohort. *Brain Behav. Immun.* Health 22.
- Fond, G., Boyer, L., Boucekine, M., Aden, L.A., Schürhoff, F., Tessier, A., et al., 2017. Validation study of the medication adherence rating scale. Results from the FACE-SZ national dataset. *Schizophr. Res.* 182, 84–89.
- Fond, G., Bulzacka, E., Boucekine, M., Schürhoff, F., Berna, F., Godin, O., et al., 2019. Machine learning for predicting psychotic relapse at 2 years in schizophrenia in the national FACE-SZ cohort. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 92, 8–18.
- George, B., Seals, S., Aban, I., 2014. Survival analysis and regression models. *J. Nucl. Cardiol. Off. Publ. Am. Soc. Nucl. Cardiol.* 21 (4), 686.
- Godin, O., Leboyer, M., Schürhoff, F., Llorca, P.M., Boyer, L., Andre, M., et al., 2018. Metabolic syndrome and illness severity predict relapse at 1-year follow-up in schizophrenia: the FACE-SZ cohort. *J. Clin. Psychiatry* 79 (6).
- Greenwood, C.J., Youssef, G.J., Letcher, P., Macdonald, J.A., Hagg, L.J., Sanson, A., et al., 2020. A comparison of penalised regression methods for informing the selection of predictive markers. *PLoS One* 15 (11), e0242730.
- Grossman, L.S., Harrow, M., Rosen, C., Faull, R., Strauss, G.P., 2008. Sex differences in schizophrenia and other psychotic disorders: a 20-year longitudinal study of psychosis and recovery. *Compr. Psychiatry* 49 (6), 523–529.
- Hui, C.L.M., Tang, J.Y.M., Leung, C.M., Wong, G.H.Y., Chang, W.C., Chan, S.K.W., et al., 2013. A 3-year retrospective cohort study of predictors of relapse in first-episode psychosis in Hong Kong. *Aust. N. Z. J. Psychiatry* 47 (8), 746–753.
- Iniesta, R., Stahl, D., McGuffin, P., 2016. Machine learning, statistical learning and the future of biological research in psychiatry. *Psychol. Med.* 46 (12), 2455–2465.
- Kuhn, M., 2008. The caret Package. *Journal of statistical package* 28 (5).
- Lähteenvuo, M., Luykx, J.J., Taipale, H., Mittendorfer-Rutz, E., Tanskanen, A., Batalla, A., et al., 2022. Associations between antipsychotic use, substance use and relapse risk in patients with schizophrenia: real-world evidence from two national cohorts. *Br. J. Psychiatry J. Ment. Sci.* 221 (6), 758–765.
- Laidi, C., Prigent, A., Plas, A., Leboyer, M., Fond, G., Chevreur, K., et al., 2018. Factors associated with direct health care costs in schizophrenia: results from the FACE-SZ French dataset. *Eur. Neuropsychopharmacol. J. Eur. Coll. Neuropsychopharmacol.* 28 (1), 24–36.
- Leboyer, M., Godin, O., Terro, E., Boukouaci, W., Lu, C.L., Andre, M., et al., 2021. Immune signatures of treatment-resistant schizophrenia: a FondaMental academic centers of expertise for schizophrenia (FACE-SZ) study. *Schizophr. Bull. Open* 2 (1).
- Lecomte, T., Potvin, S., Samson, C., Francoeur, A., Hache-Labelle, C., Gagné, S., et al., 2019. Predicting and preventing symptom onset and relapse in schizophrenia—a metareview of current empirical evidence. *J. Abnorm. Psychol.* 128 (8), 840–854.
- Leucht, S., Tardy, M., Komossa, K., Heres, S., Kissling, W., Salanti, G., et al., 2012. Antipsychotic drugs versus placebo for relapse prevention in schizophrenia: a systematic review and meta-analysis. *Lancet Lond. Engl.* 379 (9831), 2063–2071.
- Lin, C., Chen, K., Yu, J., Feng, W., Fu, W., Yang, F., et al., 2021. Relationship between TNF- $\alpha$  levels and psychiatric symptoms in first-episode drug-naïve patients with schizophrenia before and after risperidone treatment and in chronic patients. *BMC Psychiatry* 21 (1), 1–9.
- Longato, E., Vettoretti, M., Di Camillo, B., 2020. A practical perspective on the concordance index for the evaluation and selection of prognostic time-to-event models. *J. Biomed. Inform.* 108, 103496.
- Maramis, M.M., Sofyan Almahdy, M., Atika, A., Bagus Jaya Lesmana, C., Gerick Pantouw, J., 2022. The biopsychosocial-spiritual factors influencing relapse of patients with schizophrenia. *Int. J. Soc. Psychiatry* 68 (8), 1824–1833.
- Martinuzzi, E., Barbosa, S., Daoudlarian, D., Bel, W., Ali, H., Gilet, C., et al., 2019. Stratification and prediction of remission in first-episode psychosis patients: the OPTiMiSE cohort study. *Transl. Psychiatry* 9 (20).
- Menardi, G., Torelli, N., 2014. Training and assessing classification rules with imbalanced data. *Data Min. Knowl. Discov.* 28 (1), 92–122.
- Miller, B.J., Lemos, H., Schooler, N.R., Goff, D.C., Kopelowicz, A., Lauriello, J., et al., 2023. Longitudinal study of inflammation and relapse in schizophrenia. *Schizophr. Res.* 252, 88–95.
- Nieto, R.R., Carrasco, A., Corral, S., Castillo, R., Gaspar, P.A., Bustamante, M.L., et al., 2021. BDNF as a biomarker of cognition in schizophrenia/psychosis: an updated review. *Front. Psychiatry* 12.
- Pickett, K.L., Suresh, K., Campbell, K.R., Davis, S., Juarez-Colunga, E., 2021. Random survival forests for dynamic predictions of a time-to-event outcome using a longitudinal biomarker. *BMC Med. Res. Methodol.* 21 (1), 1–14.
- Pignon, B., Borel, C., Lajnef, M., Richard, J.R., Szöke, A., Hemery, F., et al., 2022. PM2.5 and PM10 air pollution peaks are associated with emergency department visits for psychotic and mood disorders. *Environ. Sci. Pollut. Res.* 29 (59), 88577–88586.
- Pillai, A., Schooler, N.R., Peter, D., Looney, S.W., Goff, D.C., Kopelowicz, A., et al., 2018. Predicting relapse in schizophrenia: is BDNF a plausible biological marker? *Schizophr. Res.* 193, 263–268.
- Porcelli, S., Bianchini, O., De Girolamo, G., Aguglia, E., Crea, L., Serretti, A., 2016. Clinical factors related to schizophrenia relapse. *Int. J. Psychiatry Clin. Pract.* 20 (2), 54–69.
- Schoretsanis, G., Kane, J.M., Correll, C.U., Rubio, J.M., 2022. Predictors of lack of relapse after random discontinuation of oral and long-acting injectable antipsychotics in clinically stabilized patients with schizophrenia: are-analysis of individual participant data. *Schizophr. Bull.* 48 (2), 296–306.
- Schürhoff, F., Fond, G., Berna, F., Bulzacka, E., Vilain, J., Capdevielle, D., et al., 2015. A national network of schizophrenia expert centres: an innovative tool to bridge the research-practice gap. *Eur. Psychiatry J. Assoc. Eur. Psychiatr.* 30 (6), 728–735.
- Schürhoff, F., Fond, G., Berna, F., Bulzacka, E., Godin, O., Boyer, L., et al., 2019. The 10-year findings from the FondaMental Academic Center of Expertise for Schizophrenia (FACE-SZ): review and recommendations for clinical practice. *L'Encephale* 45 (1), 9–14.
- Shi, X., Qu, T., Van Pottelbergh, G., van den Akker, M., De Moor, B., 2022. A resampling method to improve the prognostic model of end-stage kidney disease: abetter strategy for imbalanced data. *Front. Med.* 9.
- Siafis, S., Brandt, L., McCutcheon, R.A., Gutwinski, S., Schneider-Thoma, J., Bighelli, I., et al., 2024. Relapse in clinically stable adult patients with schizophrenia or schizoaffective disorder: evidence-based criteria derived by equiprobable linking and diagnostic test accuracy meta-analysis. *Lancet Psychiatry* 11 (1), 36–46.
- Simon, N., Friedman, J., Hastie, T., Tibshirani, R., 2011. Regularization paths for Cox's proportional hazards model via coordinate descent. *J. Stat. Softw.* 39 (5), 1–13.
- Smith, G., 2018. Step away from stepwise. *J. Big Data* 5 (1), 1–12.
- Stadtmauer, E.A., Fraietta, J.A., Davis, M.M., Cohen, A.D., Weber, K.L., Lancaster, E., et al., 2020. CRISPR-engineered T cells in patients with refractory cancer. *Science* 367 (6481).
- Tamouza, R., Krishnamoorthy, R., Giegling, I., Leboyer, M., Rujescu, D., 2020. The HLA 8.1 ancestral haplotype in schizophrenia: dual implication in neuro-synaptic pruning and autoimmunity? *Acta Psychiatr. Scand.* 141 (2), 169–171.
- Tamouza, R., Krishnamoorthy, R., Leboyer, M., 2021. Understanding the genetic contribution of the human leukocyte antigen system to common major psychiatric disorders in a world pandemic context. *Brain Behav. Immun.* 91, 731–739.
- Tiihonen, J., Mittendorfer-Rutz, E., Majak, M., Mehtälä, J., Hoti, F., Jenedien, E., et al., 2017. Real-world effectiveness of antipsychotic treatments in a nationwide cohort of 29 823 patients with schizophrenia. *JAMA Psychiatry* 74 (7), 686–693.
- Tiihonen, J., Taipale, H., Mehtälä, J., Vattulainen, P., Correll, C.U., Tanskanen, A., 2019. Association of antipsychotic polypharmacy vs monotherapy with psychiatric rehospitalization among adults with schizophrenia. *JAMA Psychiatry* 76 (5), 499–507.

van Buuren, S., Groothuis-Oudshoorn, K., 2011. Mice: multivariate imputation by chained equations in R. *J. Stat. Softw.* 45 (3), 1–67.

Vittinghoff, E., McCulloch, C.E., 2007. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am. J. Epidemiol.* 165 (6), 710–718.

Wright, M.N., Ziegler, A., 2017. Ranger: a fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw.* 77 (1), 1–17.