RESEARCH

Assessing the Role of Cytokines in Psychosis Prognosis Prediction: A Machine Learning Approach

Bart van Schie

Correspondence:
b.vanschie@students.uu.nl
Department of Science, Utrecht University, Utrecht, The Netherlands
Full list of author information is available at the end of the article

Abstract

In this study, the role of cytokines in predicting treatment outcome of first-episode psychosis (FEP) patients will be assessed.

Background: Schizophrenia is a chronic mental disorder in which early response to treatment is associated with improved prognosis. However, accurate prediction of treatment response is still a problem for modern psychiatry.

Aims: Investigate the predictive value of aggregate cytokine data in the prediction of FEP patients' clinical remission.

Methods: Data from the OPTiMiSE cohort was used to predict clinical remission as a binary outcome. Using a deep neural network, remission was predicted for patients (n=309) undergoing amisulpride treatment for 4 weeks (phase 1). In addition, remission was predicted for patients (n=57) not in remission after phase 1, who then underwent 6 weeks of either amisulpride or olanzapine treatment (phase 2).

Results: Cytokines performed better than chance in predicting treatment response for phase 1 (AUC = 0.58, 95% CI = 0.56-0.60, p = 0.024, permutation n = 1000) and phase 2 (AUC = 0.67, 95% CI = 0.59-0.75).

Conclusions: A data modality consisting of 39 cytokines performed better than chance in predicting FEP patients' clinical remission. Although these findings are modest, they suggest that cytokines should be included in a multimodal approach to predict FEP patients' treatment response.

Keywords: cytokines; schizophrenia; psychosis; prediction; OPTiMiSE; machine learning; deep learning

Introduction

Schizophrenia is a chronic [1] mental disorder that has a significant impact on the patient, its surroundings, and the society. It is characterized by comorbidity [2], high unemployment [3], and lower life expectancy [4]. The worldwide prevalence of schizophrenia is estimated at almost 1 percent [5, 6].

According to current guidelines, treatment of schizophrenia involves the use of second-generation antipsychotics whenever possible. Early response to treatment is one of the main factors associated with improved long-term prognosis [7, 8, 9, 10, 11, 12]. However, accurate prediction of treatment response is still an open problem in modern psychiatry. Psychiatrists mostly rely on a 'trial-and-error' approach for the treatment of psychiatric disorders [13, 14, 15]. Not surprisingly, a significant portion of patients do not respond well to treatment [15, 16, 17].

One way to improve treatment response is to move towards precision medicine [18, 19, 20, 21]. In general, *precision medicine* "prioritizes the individualization of care and focuses attention on unique characteristics of a particular patient" [22]. Better prediction might help to personalize treatment, potentially leading to better outcomes for the individual patient [23]. Predictions could potentially be improved by using biological variables (e.g., cytokines) in the prediction model, as they are increasingly being implicated in the etiology of schizophrenia.

Immune System Dysregulation and Schizophrenia

Although the precise pathogenesis of schizophrenia is unknown, more and more studies suggest an involvement of the immune system in schizophrenia [24, 25, 26, 27, 28]. A wide range of studies point to a dysregulation of both the innate and the adaptive immune system [29, 30, 31, 26]. It is likely that these immune changes actively contribute to clinical symptoms [32, 33, 34, 35, 36]. In this study, we will focus on cytokines, a group of small signaling molecules that coordinate both the innate and adaptive parts of the immune system.

Cytokines and Schizophrenia

Cytokines play an important role in the coordination of the immune response [37, 38]. The release of cytokines can lead to a cascade of events: attracting immune cells, production of immune cells, and the release of other cytokines [39].

Numerous studies demonstrated abnormalities in serum cytokine levels in schizophrenia patients compared to healthy controls [40, 41, 35, 42, 43, 44, 45, 46, 25, 47]. Since these studies suggest a possible link between immune dysfunction and psychosis, it was proposed that serum levels of cytokines could predict early response to treatment [48].

Cytokines and Psychosis Prognosis Prediction

Attempts to identify convincing biomarkers of treatment response in first-episode psychosis (FEP) patients have so far shown mixed results [49, 50, 51, 47, 52, 53].

One study showed that cytokines IL-6 and IL-8 could predict negative symptoms in a 6-months follow-up [50], while another demonstrated that MIP-3α serum levels were able to predict the time to remission [52]. However, these studies suffered from low statistical power issues and modest effect sizes. In a different approach, Martinuzzi et al. [54] stratified FEP patients into four clusters. For the cluster with the most severe symptoms, they found that cytokines IL-15 and CXCL12 were associated with higher odds of being non-remitters after 4 weeks. However, in a more recent study, the same research group found no cytokine biomarkers when using an unstratified approach [51].

These mixed results could be due to two reasons. First, collecting data from FEP patients requires significant time, cost, and organizational investments. Therefore, only a few studies investigated the relationship between baseline levels of cytokines and treatment response in FEP patients. Despite this challenge, FEP patients are still the preferred choice, as antipsychotic treatment itself could impact cytokine blood levels, leading to a confounding effect on the study [55, 56, 57, 58, 59]. Second, FEP patients show high heterogeneity in symptom expression and underlying

biological disease mechanisms [60, 61, 62, 63], making it more challenging to find one-size-fits-all predictors. Instead, a machine learning approach that can appreciate the complex, non-linear relations between predictors and outcomes might be more effective.

Machine Learning for Psychosis Prognosis Prediction

Nowadays, researchers are increasingly employing advanced machine learning techniques to aid in psychosis prediction for FEP patients [64, 65, 66, 21, 67]. Conventional models looking at linear relations, might not be able to capture the complex non-linear interactions between predictors and outcomes [68]. Instead, advanced (non-linear) machine learning models can potentially better detect the intricate interactions between input and output [69, 70, 68, 71, 72, 73, 74].

For example, Koutsouleris et al. [65] used a non-linear Support Vector Machine (SVM) to predict 4-week and 52-week treatment outcomes. Using baseline clinical data, they found that treatment outcomes can be reliably predicted for individual patients across multiple sites. More recently, they combined clinical data with neuro-imaging data, to predict social functioning in young patients at increased risk for psychosis. Their model outperformed clinicians, highlighting their potential in clinical practice and early intervention [66]. Even more recently, De Nijs et al. [75] used an SVM to predict three- and six-year outcomes in patients with a psychotic disorder. Using a wide range of data modalities (but not cytokines), they were able to make robust long-term prognostic predictions. This again indicates the potential of machine learning models in improving clinical judgment and decision-making. However, to date, none of these multimodal machine learning approaches included cytokines as a data modality for psychosis prognosis prediction.

Research Aims

In this study, the aim is therefore to investigate the role of aggregate cytokine data in psychosis prognosis prediction. This fills a gap in the current literature, which until now only looked at individual cytokines. These studies failed to identify convincing cytokine biomarkers of treatment response [54, 53, 76, 51]. This could be due to the high clinical and biological heterogeneity of FEP patients [60, 61, 62, 63]. Therefore, we propose to shift the focus from individual cytokines to a set of (thirty-nine) cytokines. That is, we hypothesize that using data of many cytokines together is informative in predicting FEP patients' treatment response. More specifically, it is hypothesized that baseline, aggregate data, consisting of thirty-nine cytokines, performs better than chance in FEP patients' psychosis prognosis prediction.

This hypothesis will be tested on data collected from the Optimization of Treatment and Management of Schizophrenia in Europe (OPTiMiSE) trial, a large, multimodal cohort study on FEP patients [15]. As predictors, we will use baseline data from twelve data modalities from the OPTiMiSE dataset, including cytokines. The predicted outcome is whether patients are in clinical remission. This outcome will be predicted for patients undergoing four weeks of amisulpride treatment (phase 1) and patients that failed to remit after phase 1 and then underwent six weeks of either amisulpride or olanzapine treatment (phase 2). Finally, in a *post hoc* analysis, the predictive performance of individual cytokines will be assessed, with special attention for the cytokines IL-6, IL-8, IL-18, and CXCL12, implicated as biomarkers by prior studies [50, 54].

Methods

Patients

Patients came from 27 general hospitals and clinics in 14 European countries, Israel and Australia (Clinicaltrials.gov identifier is NCT01248195). FEP patients were recruited between May 2011 and April 2016 at the participating centers from nearby healthcare facilities. Patients were 18 years and older and met DSM-IV criteria for first-episode schizophrenia, schizophreniform disorder, or schizoaffective disorder confirmed by the Mini International Neuropsychiatric Interview plus [77]. In addition, patients were within the first 2 years of onset of the first psychotic episode, with previous antipsychotic exposure of less than 15 days in the last year. A total of 495 patients signed informed consent.

Baseline Predictors and Outcome

As predictors, baseline data from twelve different modalities from the OPTiMiSE trial was used (see Table 1). These modalities were: demographics, diagnosis, lifestyle, somatic, treatment, Mini International Neuropsychiatric Interview (MINI), cytokines, Positive and Negative Syndrome Scale (PANSS), Personal Social Performance (PSP), Clinical Global Impression (CGI), Calgary Depression Scale for Schizophrenia (CDSS) and the Subjective Well-Being Under Neuroleptic Treatment Scale (SWN). There were a significant number of missing values (13% of the values were missing).

The cytokines modality consisted of thirty-nine cytokines: IL-2, IL-6, IL-7, IL-8, IL-10, IL-12p40, IL-12p70, IL-13, IL-15, IL-16, IL-17, IL-18, IL-21, IL-23, IL-27, IFN-γ, chemokines C-C motif ligand (CCL)-2, CCL3, CCL4, CCL11, CCL13, CCL17, CCL19, CCL20, CCL22, CCL26, CCL27, C-X3-C motif chemokine ligand (CX3CL)-1, CXCL10, CXCL11, CXCL12, TNF-α, TNF-β, vascular endothelial growth factor (VEGF), C reactive protein (CRP), serum amyloid A (SAA), soluble intercellular adhesion molecule-1 (sICAM-1) and soluble vascular adhesion molecule-1 (sVCAM-1).

The outcome variable was symptomatic remission based on the criteria of Andreasen et al. [78]: a score of ≤ 3 (on a scale ranging from 1 to 7) simultaneously on all eight PANSS items (P1, P2, P3, N1, N4, N6, G5 and G9).

Study Design

Patients' symptomatic remission was predicted for phase 1 and phase 2 of the OPTiMiSE trial and for two strategies to handle missing data - imputation and neutralization.

Phase 1 (4 Weeks) and Phase 2 (10 Weeks)

From the study onset, all patients were treated for 4 weeks with the antipsychotic amisulpride in an open-label design (phase 1). Subsequently, patients who did not meet clinical remission at 4 weeks were randomly assigned to continue on amisulpride or to switch to olanzapine during a 6-week double-blind trial (phase 2).

Table 1 Data Modalities of the OPTiMiSE Dataset.

	Modality	No. Features	Missing values	Description
1	Demographics	14	6%	Socio-demographic features
2	Diagnosis	7	6%	Illness related features
3	Lifestyle	7	21%	Use of substances like drugs and alcohol
4	Somatic	11	13%	Physical examination
5	Treatment	1	14%	Average dosage of medication
6	MINI	67	8%	Psychiatric comorbidity
7	Cytokines	39	20%	Small proteins important in cell signaling
8	PANSS	30	8%	Positive and Negative Syndrome Scale
9	PSP	5	11%	Personal and Social Performance Scale
10	CGI	1	9%	Clinical Global Impression
11	CDSS	9	11%	Measurement scale about depression
12	SWN	20	16%	Illness related features
	Total	**211**	**13%**	-

Handling Missing Values: Imputation and Neutralization

Clinical trials can suffer from a considerable amount of missing values [79, 80]. As seen in Table 1, the OPTiMiSE trial is no exception. Missing values can potentially lead to a bias in the results [81]. For example, dropout in clinical trials of antipsychotic medication could be related to symptom severity [82]. Therefore, it is important to check for such bias, especially concerning important outcome variables. To provide such a (quick) check, see Figure 5 in Appendix A.1 for the differences in total PANSS score, between patients with and without complete data.

Moreover, to minimize the effect of any potential bias, two different strategies to handle missing values were used. This can be seen as a form of lightweight sensitivity analysis, following missing values guidelines in clinical research [83]. Specifically, FEP patients' clinical remission was predicted under two distinct strategies to handle missing values - k-Nearest Neighbors (k-NN) data imputation and neutralization [84].

Data Preprocessing

Dropping Data

The data preprocessing pipeline started with a total of 495 patients that started the OPTiMiSE trial and signed informed consent. Patients that did not have any measurement for one of the modalities were dropped, reducing the sample size to 383 patients. Next, for the 4-week analysis, patients were dropped that did complete phase 1 (resulting n=309). Similarly, for the 10-week analysis patients not finishing phase 2 were dropped (resulting n=57).

Handling Missing Data

Two distinct strategies were applied to handle missing data - k-NN and neutralization [84]. With k-NN, missing values were imputed using the `k-nearest neighbor` library from `scikit-learn` [85]. In the neutralization strategy, missing values were neutralized using an experimental fully-connected layer. This layer replaces the

fully-connected `Dense` layer and neutralizes the effect of missing values (and does nothing otherwise). The performance of this imputation technique is similar to other imputation techniques like k-NN, mean imputation, and zero imputation [84].

Preparing Data

The data modalities contained binary, categorical, and continuous features. For the binary features 0/1 encoding was used and for the categorical features one-hot encoding. All continuous features were standardized before feeding them to the model. Some cytokines contained a small amount of lower limit of detection (LLOD) values. These LLOD's were replaced with the minimum value of that cytokine, divided by two.

Neural Network Architecture, Training and Testing

Architecture

Using the architecture of Kia et al. [84], a multimodal neural network was employed. As shown in Figure 1, the model has one input layer and four fully-connected layers: representation learning, modality-specific classification, fusion, and classification. In the input layer, the model receives data from twelve different data modalities. To accommodate for the neutralization of missing values, `nanDense` [84] layers were positioned right after the input layer. These layers can neutralize the effect of missing values and behave as a normal `Dense` layer otherwise. To determine the number of neurons in the representation learning layer, a simple heuristic was employed: the number of neurons for a given modality equals the number of features of that modality. After representation learning, modality-specific classification layers classified each modalities' representation into the outcome variable. These layers consisted of two `Softmax` neurons, representing remission/non-remission. Next, these classifications were merged with the representation learning and fed into a fusion layer of five neurons. Finally, two Softmax neurons were used to classify the fused data into remission/non-remission. Dropout layers with a probability of 0.1 were applied before any fully-connected layer. Both the representation learning and the fusion layer used a `ReLU` [86] activation function. As an optimization algorithm, an `Adam` [87] optimizer (learning rate: 0.0003) was used to minimize the categorical cross-entropy loss in the output layer.

Training and Testing

The model was trained and tested in ten identical runs (for implementation details see Appendix A.2). For each run, 10-fold cross-validation was used to evaluate the performance of the model. To deal with the unbalanced distribution of the outcome variable (67% of patients get remitted), `sklearn`'s `Stratified K-Folds cross-validator` was used. This cross-validator is a variation of KFold that returns stratified folds. The folds are made by preserving the percentage of samples for each class, in each fold of K-fold cross-validation. In each run, the model was trained on 9 of the folds for 50 epochs, and tested on the left-out fold, until all folds were tested on.

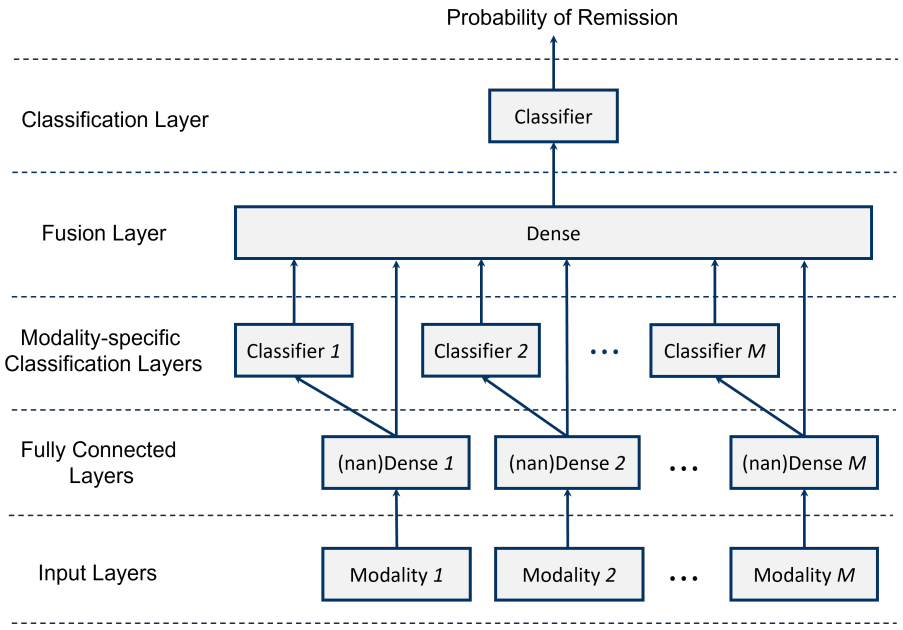

Figure 1 Diagram of the Deep Neural Network Used for Psychosis Prognosis Prediction.

Counterfactual Predictions of Individual Cytokines for Phase 1

For phase 1 predictions, counterfactual interpretation (CFI) was applied to assess the predictive value of individual cytokines in psychosis prognosis prediction. CFI provides a tool for interpreting complex AI models [88]. It describes a causal situation in the form: "What would have happened to outcome Y, if everything stayed the same, except for some factor X?". To apply this to a deep neural network, a special fully-connected layer was used that can neutralize the effect of missing values [84]. Using this layer, it was possible to 'turn off' individual features (e.g., cytokines) and check whether the prediction improved (or worsened).

Regarding training and testing the model, all the previous steps were taken, with one exception. Instead of training the model on all (twelve) modalities, it was trained on only the cytokine modality. At the end of each run, CFI was introduced in the following way: Every cytokine was, iteratively, 'turned off', by converting its values to missing values (i.e., `NaN`). This way, the input consisted of 38 'turned on' cytokines and one neutralized 'turned off' cytokine. Subsequently, patients' clinical remission was predicted with this adapted input layer.

Statistical Analysis

AUC, Sensitivity, and Specificity

In each run, the predictive performance of the model was evaluated by computing the area under the receiver-operating curve (AUC). An AUC of 1.0 reflects perfect performance and 0.5 chance level performance. After ten runs, AUC scores were averaged, and a 95% confidence interval (CI) was computed. To compute the AUC, the `roc_auc_score` function from `scikit-learn` was used [85]. Remember that the classification task assigns probabilities to either remission or non-remission. AUC is especially suited for these kinds of tasks, as it can analyze the prediction

more 'deeply'. AUC represents the probability that a random positive example (i.e., remission) is ranked higher than a random negative example (i.e., non-remission). Unlike accuracy, AUC considers all possible decision thresholds (not only the default 0.5) and because of this can provide a 'broader' view of the performance of the classifier. Finally, sensitivity and specificity were measured, to assess how well (or biased) the model predicts both remission and non-remission.

Permutation Analysis

Permutation testing [89, 90] was applied if the AUC score of the cytokine modality was too close to chance level (cut-off point: AUC $\leq$ 0.6). To get a chance level null-distribution, labels were randomly permuted 1000 times. As a result, a null-distribution of a thousand AUC scores was obtained, based on the null hypothesis that the model performed on chance-level. Subsequently, actual AUC scores were acquired, by feeding the model untouched data (i.e., non-shuffled labels). To obtain the p-value, the proportion of AUC values in the null-distribution, greater or equal than the observed AUC value, was computed.

Computing Individual Cytokines' Contribution to Prediction Certainty

For the *post hoc* analysis, individual cytokines' predictive performance was computed by checking whether the prediction (for a given cytokine) came closer to the true label (i.e., 0 or 1) or not.

For example, suppose the true label of a patient was 1 (i.e., remission), and the prediction with a given cytokine 'turned off' was 0.7. Then, the resulting prediction uncertainty would be 0.3. This would then be compared to the uncertainty with the given cytokine 'turned on'. If the latter uncertainty was 0.1, it implied that the prediction became 0.2 (i.e., 20%) more certain because of that cytokine.

Finally, these individual cytokines' contributions were analyzed for the full sample and the special cluster found by Martinuzzi et al. [54].

Results

Socio-demographic and Clinical Characteristics of Patients

Table 2 summarizes the clinical and socio-demographic characteristics of the patients included in the analysis.

The sample of patients that finished phase 1 was composed of 309 patients. The mean age was 26 years and 71% of them were male. Of the patients not in remittance after phase 1 (n=95), 76% were male, 32% overweight and 66% unemployed. Furthermore, the total PANSS score of non-remitted patients (84.1) was higher than that of remitted patients (74.4).

A total of 57 patients completed phase 2, of which 27 remitted and 30 did not remit. Among others, non-remitted patients were more often overweight (37%), older (27 years), more recreational drug use (47%), and were more often living alone (27%) than remitted ones.

Prognosis Prediction for Patients Completing Phase 1 (4 Weeks)

Looking at Figure 2, AUC scores suggest that the cytokines modality was among the best-performing modalities, with an AUC of 0.58 (95% CI = 0.56-0.60, p=0.024,

Table 2 Socio-demographic and Clinical Details of Patients at Week 4 and 10. Data are shown as mean (standard deviation, SD), or *n* (%). Recreational drug use is defined as having used recreational drugs at least once in a lifetime. Abbreviations: BMI, body mass index; CDSS, Calgary Depression Scale for Schizophrenia; CGI, clinical global impression; PANSS, Positive and Negative Syndrome Scale.

	Week 4		**Week 10**	
	Remitted	**Non-remitted**	**Remitted**	**Non-remitted**
Number of patients	214	95	27	30
Age (years)	26.8 (6.3)	25.6 (5.6)	23.8 (4.0)	27.0 (6.8)
Male	69%	76%	70%	77%
BMI (kg/m^2)	23.2 (4.0)	23.4 (4.4)	23.1 (5.3)	22.9 (4.1)
Waist circumference (cm)	82.7 (11.5)	84.6 (11.8)	82.9 (10.9)	84.1 (12.3)
Overweight (BMI ≥ 25)	24%	32%	19%	30%
Employed/Student	48%	34%	22%	37%
Recreational drug use	50%	45%	33%	47%
Living alone	17%	21%	7%	27%
PANSS (total)	74.4 (19.4)	84.1 (16.6)	84.8 (15.1)	89.1 (17.4)
CDSS depression score	13.1 (4.5)	13.5 (4.5)	12.2 (4.4)	14.5 (4.0)
CGI severity	5.4 (1.0)	5.6 (0.9)	5.5 (1.0)	5.7 (1.0)

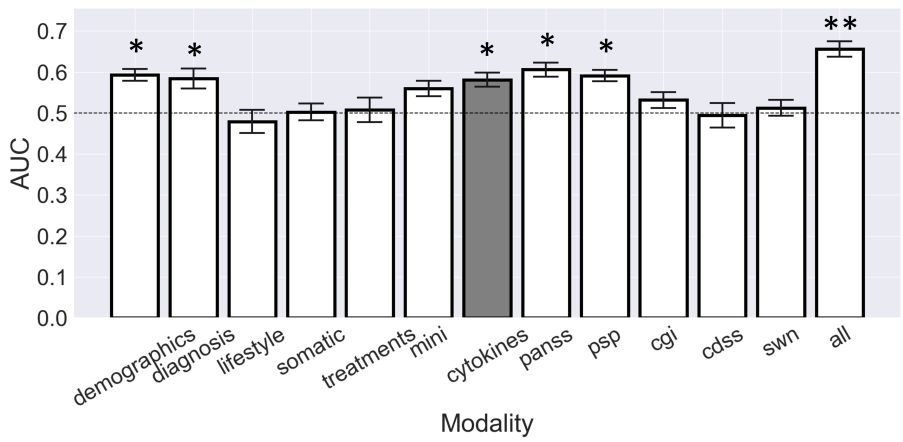

Figure 2 AUC Scores for Predicting Remission After Phase 1 (4 Weeks). The figure shows the mean AUC and SD error bar, after training and testing the model for ten runs. Included are the AUC scores for each of the twelve modalities in the OPTiMiSE trial.
*$p \leq 0.05$. **$p \leq 0.01$.

permutation n=1000). In Table 3, we can see the classification performance of all modalities. Nearly half of the data modalities provided significantly better than chance AUC scores, namely demographics, diagnosis, cytokines, PANSS, and PSP. Conversely, the MINI, lifestyle, somatic, treatments, CGI, CDSS, and SWN modalities did not perform better than chance.

It is interesting to note that models trained on cytokines and PANSS, had relatively high sensitivity and low specificity. This means that these models were especially successful at predicting remission but much less at predicting non-remission.

Demographics and diagnosis were also among the high performers but had a much more moderate difference between sensitivity and specificity.

Furthermore, the use of either neutralization or imputation to handle missing values seemed to have little effect, although AUC scores are somewhat higher for each modality in the neutralization condition. See Table 6 in Appendix A.3 for the classification performance between both strategies to handle missing values.

Table 3 Classification Performance for FEP Patients Completing Phase 1 (Week 4). Data are mean AUC (SD), sensitivity, and specificity scores. Significance levels were confirmed by permutation testing (n=1000). *$p \leq 0.05$. **$p \leq 0.01$.

Modality	AUC	Sensitivity	Specificity	p-value
Demographics	0.59 (0.02)	0.70	0.40	0.019*
Diagnosis	0.58 (0.02)	0.73	0.37	0.022*
Lifestyle	0.49 (0.03)	0.64	0.33	0.574
Somatic	0.49 (0.03)	0.69	0.30	0.526
Treatments	0.52 (0.03)	0.65	0.38	0.231
MINI	0.56 (0.02)	0.72	0.38	0.079
Cytokines	0.58 (0.02)	0.74	0.32	0.024*
PANSS	0.60 (0.02)	0.72	0.38	0.015*
PSP	0.58 (0.02)	0.67	0.42	0.013*
CGI	0.54 (0.03)	0.60	0.45	0.079
CDSS	0.50 (0.03)	0.73	0.30	0.491
SWN	0.51 (0.03)	0.70	0.31	0.444
All	0.64 (0.02)	0.85	0.31	0.001**

Prognosis Prediction for Patients Completing Phase 2 (10 Weeks)

The cytokine modality performed best in predicting FEP patients completing phase 2, with an AUC of 0.66 (95% CI = 0.61-0.71). Interestingly, the performance is even better than with all modalities fused together. Furthermore, the difference in modalities' AUC scores between imputation and neutralization was well within their respective confidence intervals. This suggests that the models' behavior was mostly invariant to either strategy to handle missing values (see Table 7 in Appendix A.4). Interestingly, in contrast to the phase 1 results, there was not much of a difference between sensitivity and specificity for the cytokine modality.

Individual Cytokines' Contribution to Prediction Certainty

A *post hoc* analysis of individual cytokines' predictive performance suggests that cytokine IL-18 contributes most to the certainty of the prediction. In this unstratified approach, cytokines IL-6, IL-8, IL-15, CXCL12 performed mediocre to poor. See Figure 5 in Appendix A.5 for the difference between these cytokines and the top 5 cytokines.

In addition, a *post hoc* test was performed for a cluster of 97 FEP patients. For this cluster, cytokines CCL22, IL-15, and IL-12p40 show the biggest improvement in prediction certainty. Figure 5 in Appendix A.6 shows the top 10 performing cytokines.

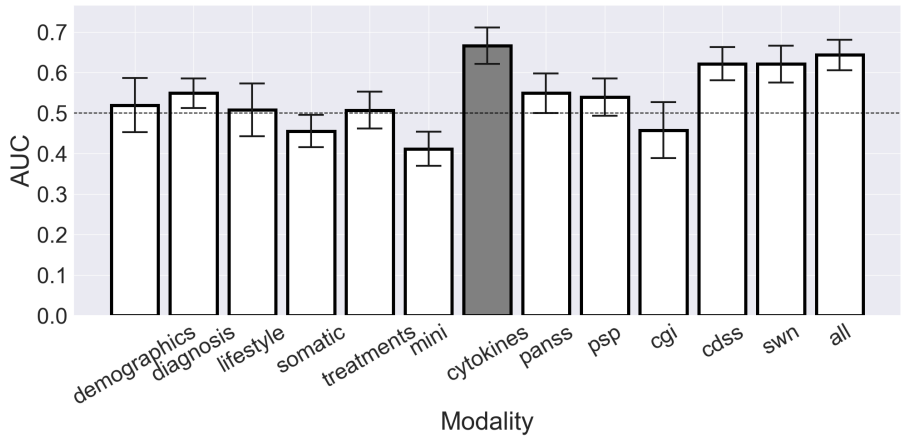

Figure 3 AUC Scores for Predicting Remission After Phase 2 (10 Weeks). The figure shows the mean AUC and SD error bar, after training and testing the model for ten runs. Included are the AUC scores for each of the twelve modalities in the OPTiMiSE trial.

Table 4 Classification Performance for FEP Patients Completing Phase 2 (Week 10). Data are mean AUC (SD), sensitivity, and specificity scores. Significance levels were not confirmed by permutation testing as the AUC for cytokines was greater than the cut-off point of 0.6

Modality	AUC	Sensitivity	Specificity
Demographics	0.54 (0.05)	0.57	0.54
Diagnosis	0.49 (0.05)	0.48	0.50
Lifestyle	0.55 (0.07)	0.56	0.53
Somatic	0.41 (0.06)	0.42	0.48
Treatments	0.48 (0.06)	0.53	0.42
MINI	0.43 (0.05)	0.44	0.43
Cytokines	0.66 (0.05)	0.60	0.61
PANSS	0.51 (0.05)	0.48	0.52
PSP	0.52 (0.08)	0.53	0.48
CGI	0.49 (0.06)	0.51	0.49
CDSS	0.62 (0.04)	0.67	0.57
SWN	0.60 (0.05)	0.53	0.62
All	0.63 (0.04)	0.54	0.69

Discussion

The aim of this study was to assess whether aggregate baseline cytokine data could improve psychosis prognosis prediction. The results suggest this is indeed the case. Baseline cytokine levels show better than chance performance in predicting treatment response for FEP patients completing phase 1 (4 weeks) of the OPTiMiSE trial. Even a stronger performance was found for patients finishing phase 2 (10 weeks).

Both results were invariant under two distinct strategies to handle missing values. In addition, significant predictive value was found for other data modalities, namely

demographics, diagnosis, PANSS, and PSP, but not for MINI, lifestyle, somatic, treatments, CGI, CDSS, and SWN.

To date, no other study explored aggregate cytokine data in relation to predicting treatment response in drug-naive FEP patients. This study fills that gap and shows a real promise for cytokines as an aggregate input modality in psychosis prognosis prediction.

Our findings have several interesting practical implications. First, it provides support for the continued inclusion of cytokines in multimodal cohort studies for schizophrenia patients. In addition, it may guide decision-making on whether to include other predictive modalities. After all, about half of the data modalities did not perform better than chance. Second, it may sway research groups, aiming to develop clinically relevant prediction models, to include cytokines in their set of predictors. So far, few of such efforts have included cytokines, but the findings from this study might convince researchers to include them by default.

Finally, a *post hoc* analysis was performed to investigate individual biomarkers implicated by prior research [50, 54]. In this research, He et al. [50] found that cytokines IL-6 and IL-8 predicted negative symptoms, while Martinuzzi et al. [54], identified IL-15 and CXCL12 as useful biomarkers in a cluster of FEP patients. Our findings do not support the predictive value of IL-6 and IL-8. However, our analysis predicted 4-week remission, while IL-6 and IL-12 were identified in a 6-month follow-up. Conversely, the results do support the predictive value of cytokine IL-15 in a special cluster of patients. This corroborates the result of Martinuzzi et al. [54], who also predicted 4-week remission. Nonetheless, the findings from our *post hoc* analysis should be met with caution. The effect sizes were relatively small and the standard deviations relatively high. In addition, we employed an experimental method, by applying CFI to a deep neural network architecture. Ideally, this method first needs to be further validated by future research.

Going forward, we have shown that several data modalities (including cytokines) have value in psychosis prognosis prediction. This was already apparent in the current state of research [65, 66, 75] and if anything, has only been corroborated by this study's results. Looking at the future, it is this multimodal approach that might bring us closer to precision medicine.

Future Directions

Precision Medicine

A growing body of literature shows a real promise for a movement towards precision medicine [21]. Applying advanced machine learning to the likes of imaging, clinical, cognitive, and biological data can potentially improve the prediction of psychosis outcomes. This, in turn, can assist clinicians to better tailor (i.e., individualize) treatment, which can potentially lead to better outcomes for the individual patient [23]. One way to achieve such (effective) precision medicine, could be a growing embracement of implementation research.

Implementation Research

A growing shift towards implementation research could improve the clinical relevance of prediction models and provide these models with (much) more data.

In a review of individualized prediction models for clinical practice, it was found that only 0.2% of these studies could be classified as an implementation study [19]. In this review, the authors state that advancements in precision medicine could be held back by limited replication and lack of implementation research in real-world clinical practice.

One promise of implementation research is that it could potentially solve the problem of external validation [91, 68]. That is, many models are (too much) tuned to the intricacies of the data they are trained on and have difficulties achieving the same performance on unseen data. Conversely, with implementation research, these models can be trained and tested on the very population they are meant for. This way, the problem of external validation in clinical prediction models could largely be solved. Note, however, that samples from implementation research will likely have a higher heterogeneity of patients. After all, every type of patient will enter clinical practice, while not every type may enter clinical trials. Such an increased heterogeneity of the sample can decrease the performance of the model [92].

Another promise of implementation research is that it could open up a massive stream of clinically relevant data. Through implementation research, data collection can be embedded in day-to-day clinical practice. Here, data should be ideally collectable using methods that are widely available, do not require an excessive amount of patient or clinician time, and that have a reasonable cost [21]. When applied on a large scale, this could lead to much larger sample sizes. This, in turn, could allow data-driven machine learning models to achieve unmatched performance. In addition, the individual samples themselves can become much more data-rich. After all, if data is collected at every visit, it increasingly becomes more longitudinal in nature. With time, patients' longitudinal data can include a wide range of interventions (e.g., shifting antipsychotics, treatments, lifestyles). Such data could potentially enable machine learning algorithms to learn which intervention works best, for what type of patient (i.e., precision medicine). Finally, when these prediction models reach a high enough performance, exciting possibilities in the realm of explainable Artificial Intelligence (AI) arise.

Explainable AI

One such opportunity is counterfactual interpretation (CFI). Applying CFI on (future) high-performing prediction models can make it possible to quantify the contribution of each unique feature used in the prediction. Such a development could provide new insights in the underlying mechanisms of syndromes like schizophrenia. In addition, CFI can provide easy and individualized therapeutic targets for clinicians. Any factor that can be conveniently manipulated (e.g., lifestyle changes, stress levels, cytokines), can suddenly become a valuable treatment option. For example, there are already drugs with anti-inflammatory or immune-modulating properties that can manipulate the levels of specific cytokines [93, 94].

Explainable AI with CFI could also have a more practical, short-term, benefit. As seen in the *post hoc* analysis, CFI could shed light on the relative importance of individual cytokines in psychosis prognosis prediction. This way, it might be possible to be more selective when deciding on which cytokines to include in clinical trials and prediction models.

Strengths and Limitations

This study had several limitations. First, although the predictive performance of the cytokine modality was better than chance, it was still modest. Note, however, that it was not this study's aim to design and tune a model for maximum performance. Second, the cytokine performance for phase 1 showed high sensitivity and low specificity, indicating a bias towards predicting remission. This could be due to the unbalanced class distribution and can potentially be solved by oversampling the minority class. However, oversampling can also worsen model performance [95]. Third, there was a lack of external validation of the results. Ideally, the model would have been trained and evaluated on similar datasets, coming from different samples. Unfortunately, it is difficult to access comparable datasets, especially when dealing with FEP patients. Fourth, the sample size for phase 2 was modest (n=57), as high remission and dropout rates reduced the number of samples. Because of this relatively small sample size, the phase 2 results might have been due to (overfitting) a homogeneous sample [92] and might not generalize well to other samples.

There were also some limitations concerning the OPTiMiSE trial. First, predictions were only made for patients treated with amisulpride (phase 1) or amisulpride and olanzapine (phase 2). This might limit the generalizability of the results. After all, the results could have been different if patients had been treated with other popular antipsychotics [96] like clozapine and risperidone. Second, some patients dropped out of the study during the OPTiMiSE trial. This may limit the representativeness of the patients used in prediction. Third, selection bias may be prevalent as patients included in OPTiMiSE may differ from those that did not consent to participation. Fourth, although the OPTiMiSE trial is one of the largest cohorts with FEP patients to date, the sample size may not have been sufficient to appreciate the heterogeneity of out-of-study schizophrenia patients. Increasing the sample size will likely improve generalizability, by including a wider range of clinical and biological representations of schizophrenia. However, as indicated, increasing the heterogeneity of the sample can also decrease performance [92].

This study also had several strengths. First, data came from a relatively large sample of drug-naive FEP patients. This diminishes the effect of possible confounding factors, like illness duration and long exposure to antipsychotics [57]. For example, the latter can have a direct effect on cytokine levels themselves [56]. Second, the OPTiMiSE trial being a multi-center European cohort, potentially improves the generalizability of the results. After all, patients came from 27 centers, from 14 European countries, Israel and Australia. Third, the longitudinal nature of the data made it possible to perform psychosis prognosis prediction for different points in time (e.g., 4 and 10 weeks). This is informative, as phase 2 patients might be clinically different (i.e., more chronic) than phase 1 patients. Fourth, a multimodal set of predictors was used, making it possible to assess the performance of each of these modalities individually. This can be important in deciding which modalities to include in future clinical studies and prediction models. Fifth, two different strategies to handle missing values were applied. This can reduce the potential bias coming from replacing (i.e., imputing) missing values. Sixth, this study tried to avoid publication bias [97], by taking the average of many results. This way, a form of 'cherry-picking' the best results was avoided. Seventh, although outside the scope

of this study, a time-consuming permutation test was applied to test whether the cytokine modality performed better than chance. This test adds to the robustness of the results, as it makes clear that the results were not due to an inherent bias in the model.

Conclusions

A deep multimodal neural network was applied to assess the value of cytokines in psychosis prognosis prediction. Using data from first-episode psychosis (FEP) patients in the OPTiMiSE trial, the predictive performance of twelve modalities (including cytokines) was assessed. Treatment response was predicted for patients completing phase 1 (4 weeks) and phase 2 (10 weeks) of the OPTiMiSE trial. The results show that the cytokine data modality offers modest but better-than-chance predictive value for patients completing phase 1. The predictive performance for phase 2 is even better and outperforms all the other modalities. To date, no other study assessed the role of aggregate cytokine data in FEP patients' psychosis prognosis prediction. This study shows that cytokines (and other modalities) hold a genuine predictive promise. These findings suggest cytokines should be included in a multimodal approach towards psychosis prognosis prediction. Looking at the future, a growing shift towards implementation research might lead to improvements in prediction models and data collection. Such a shift could bring us closer to precision medicine, by increasingly enabling machine learning models to decide which intervention is best for what type of patient.

Acknowledgments
I thank the patients and their families who participated in the OPTiMiSE trial. Furthermore, I would like to thank the members of the OPTiMiSE study group for their access to the data. In addition, I want to thank Cyprien Gillet and colleagues for sharing their cytokines data and making code available. Last but not least, I would like to thank dr. Hugo Schnack and dr. Seyed Mostafa Kia for supervising me during this thesis project.

Code availability
All code and scripts can be made available upon request.

Ethics declarations
Conflict of Interest
The authors declare that they have no conflict of interest.

Competing interests
The authors declare that they have no competing interests.

Author details
Department of Science, Utrecht University, Utrecht, The Netherlands.

References

1. Barbato, A.: Psychiatry in transition: outcomes of mental health policy shift in italy. Australian and New Zealand Journal of Psychiatry **32**(5), 673–679 (1998)
2. Goff, D.C., Cather, C., Evins, A.E., Henderson, D.C., Freudenreich, O., Copeland, P.M., Bierer, M., Duckworth, K., Sacks, F.M.: Medical morbidity and mortality in schizophrenia: guidelines for psychiatrists. Journal of Clinical Psychiatry **66**(2), 183–194 (2005)
3. Marwaha, S., Johnson, S.: Schizophrenia and employment. Social psychiatry and psychiatric epidemiology **39**(5), 337–349 (2004)
4. Chesney, E., Goodwin, G.M., Fazel, S.: Risks of all-cause and suicide mortality in mental disorders: a meta-review. World psychiatry **13**(2), 153–160 (2014)
5. Saha, S., Chant, D., Welham, J., McGrath, J.: A systematic review of the prevalence of schizophrenia. PLoS medicine **2**(5), 141 (2005)
6. Moreno-Küstner, B., Martin, C., Pastor, L.: Prevalence of psychotic disorders and its association with methodological issues. a systematic review and meta-analyses. PloS one **13**(4), 0195687 (2018)
7. Leucht, S., Arbter, D., Engel, R.R., Kissling, W., Davis, J.: How effective are second-generation antipsychotic drugs? a meta-analysis of placebo-controlled trials. Molecular psychiatry **14**(4), 429–447 (2009)
8. Lieberman, J.A., Dixon, L.B., Goldman, H.H.: Early detection and intervention in schizophrenia: a new therapeutic model. Jama **310**(7), 689–690 (2013)

9. Zhang, J.-P., Gallego, J.A., Robinson, D.G., Malhotra, A.K., Kane, J.M., Correll, C.U.: Efficacy and safety of individual second-generation vs. first-generation antipsychotics in first-episode psychosis: a systematic review and meta-analysis. International Journal of Neuropsychopharmacology **16**(6), 1205–1218 (2013)
10. Kane, J.M., Robinson, D.G., Schooler, N.R., Mueser, K.T., Penn, D.L., Rosenheck, R.A., Addington, J., Brunette, M.F., Correll, C.U., Estroff, S.E., *et al.*: Comprehensive versus usual community care for first-episode psychosis: 2-year outcomes from the nimh raise early treatment program. American Journal of Psychiatry **173**(4), 362–372 (2016)
11. Cheng, S.C., Schepp, K.G.: Early intervention in schizophrenia: a literature review. Archives of psychiatric nursing **30**(6), 774–781 (2016)
12. Fusar-Poli, P., Díaz-Caneja, C.M., Patel, R., Valmaggia, L., Byrne, M., Garety, P., Shetty, H., Broadbent, M., Stewart, R., McGuire, P.: Services for people at high risk improve outcomes in patients with first episode psychosis. Acta Psychiatrica Scandinavica **133**(1), 76–85 (2016)
13. Bzdok, D., Meyer-Lindenberg, A.: Machine learning for precision psychiatry: opportunities and challenges. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging **3**(3), 223–230 (2018)
14. Arango, C., Kapur, S., Kahn, R.S.: Going beyond "trial-and-error" in psychiatric treatments: OPTiMiSE-ing the treatment of first episode of schizophrenia. Oxford University Press US (2015)
15. Kahn, R.S., van Rossum, I.W., Leucht, S., McGuire, P., Lewis, S.W., Leboyer, M., Arango, C., Dazzan, P., Drake, R., Heres, S., *et al.*: Amisulpride and olanzapine followed by open-label treatment with clozapine in first-episode schizophrenia and schizophreniform disorder (optimise): a three-phase switching study. The Lancet Psychiatry **5**(10), 797–807 (2018)
16. Barnes, T.R., Drake, R., Paton, C., Cooper, S.J., Deakin, B., Ferrier, I.N., Gregory, C.J., Haddad, P.M., Howes, O.D., Jones, I., *et al.*: Evidence-based guidelines for the pharmacological treatment of schizophrenia: updated recommendations from the british association for psychopharmacology. Journal of Psychopharmacology **34**(1), 3–78 (2020)
17. Potkin, S.G., Kane, J.M., Correll, C.U., Lindenmayer, J.-P., Agid, O., Marder, S.R., Olfson, M., Howes, O.D.: The neurobiology of treatment-resistant schizophrenia: paths to antipsychotic resistance and a roadmap for future research. NPJ schizophrenia **6**(1), 1–10 (2020)
18. Manchia, M., Pisanu, C., Squassina, A., Carpiniello, B.: Challenges and future prospects of precision medicine in psychiatry. Pharmacogenomics and personalized medicine **13**, 127 (2020)
19. Salazar de Pablo, G., Studerus, E., Vaquerizo-Serrano, J., Irving, J., Catalan, A., Oliver, D., Baldwin, H., Danese, A., Fazel, S., Steyerberg, E.W., *et al.*: Implementing precision psychiatry: a systematic review of individualized prediction models for clinical practice. Schizophrenia bulletin **47**(2), 284–297 (2021)
20. Pedrini, M., Cao, B., Nani, J.V.S., Cerqueira, R.O., Mansur, R.B., Tasic, L., Hayashi, M.A., McIntyre, R.S., Brietzke, E.: Advances and challenges in development of precision psychiatry through clinical metabolomics on mood and psychotic disorders. Progress in Neuro-Psychopharmacology and Biological Psychiatry **93**, 182–188 (2019)
21. Tognin, S., van Hell, H.H., Merritt, K., Winter-van Rossum, I., Bossong, M.G., Kempton, M.J., Modinos, G., Fusar-Poli, P., Mechelli, A., Dazzan, P., *et al.*: Towards precision medicine in psychosis: benefits and challenges of multimodal multicenter studies—psyscan: translating neuroimaging findings from research into clinical practice. Schizophrenia bulletin **46**(2), 432–441 (2020)
22. Tonelli, M.R., Shirts, B.H.: Knowledge for precision medicine: mechanistic reasoning and methodological pluralism. Jama **318**(17), 1649–1650 (2017)
23. Marwaha, S., Thompson, A., Upthegrove, R., Broome, M.R.: Fifteen years on–early intervention for a new generation. The British Journal of Psychiatry **209**(3), 186–188 (2016)
24. Bishop, J.R., Zhang, L., Lizano, P.: Inflammation subtypes and translating inflammation-related genetic findings in schizophrenia and related psychoses: a perspective on pathways for treatment stratification and novel therapies. Harvard review of psychiatry **30**(1), 59 (2022)
25. Goldsmith, D., Rapaport, M., Miller, B.: A meta-analysis of blood cytokine network alterations in psychiatric patients: comparisons between schizophrenia, bipolar disorder and depression. Molecular psychiatry **21**(12), 1696–1709 (2016)
26. Leirer, D.J., Iyegbe, C.O., Di Forti, M., Patel, H., Carra, E., Fraietta, S., Colizzi, M., Mondelli, V., Quattrone, D., Lally, J., *et al.*: Differential gene expression analysis in blood of first episode psychosis patients. Schizophrenia research **209**, 88–97 (2019)
27. Horváth, S., Mirnics, K.: Immune system disturbances in schizophrenia. Biological psychiatry **75**(4), 316–323 (2014)
28. Miller, B.J., Goldsmith, D.R.: Evaluating the hypothesis that schizophrenia is an inflammatory disorder. Focus **18**(4), 391–401 (2020)
29. Brown, A.S., Derkits, E.J.: Prenatal infection and schizophrenia: a review of epidemiologic and translational studies. American Journal of Psychiatry **167**(3), 261–280 (2010)
30. Drexhage, R.C., Knijff, E.M., Padmos, R.C., Heul-Nieuwenhuijzen, L.v.d., Beumer, W., Versnel, M.A., Drexhage, H.A.: The mononuclear phagocyte system and its cytokine inflammatory networks in schizophrenia and bipolar disorder. Expert review of neurotherapeutics **10**(1), 59–76 (2010)
31. Bergink, V., Burgerhout, K.M., Weigelt, K., Pop, V.J., de Wit, H., Drexhage, R.C., Kushner, S.A., Drexhage, H.A.: Immune system dysregulation in first-onset postpartum psychosis. Biological psychiatry **73**(10), 1000–1007 (2013)
32. Michel, M., Schmidt, M.J., Mirnics, K.: Immune system gene dysregulation in autism and schizophrenia. Developmental neurobiology **72**(10), 1277–1287 (2012)
33. Mueller, N.: Cox-2 inhibitors as antidepressants and antipsychotics: clinical evidence. Curr Opin Investig Drugs **11**(1), 31–42 (2010)
34. Müller, N., Weidinger, E., Leitner, B., Schwarz, M.J.: The role of inflammation in schizophrenia. Frontiers in neuroscience **9**, 372 (2015)
35. Upthegrove, R., Manzanares-Teson, N., Barnes, N.M.: Cytokine function in medication-naive first episode

psychosis: a systematic review and meta-analysis. Schizophrenia research **155**(1-3), 101–108 (2014)
36. Khandaker, G.M., Cousins, L., Deakin, J., Lennox, B.R., Yolken, R., Jones, P.B.: Inflammation and immunity in schizophrenia: implications for pathophysiology and treatment. The Lancet Psychiatry **2**(3), 258–270 (2015)
37. Takeuchi, O., Akira, S.: Pattern recognition receptors and inflammation. Cell **140**(6), 805–820 (2010)
38. Miller, B.J., Goldsmith, D.R.: Towards an immunophenotype of schizophrenia: progress, potential mechanisms, and future directions. Neuropsychopharmacology **42**(1), 299–317 (2017)
39. Schaper, F., Rose-John, S.: Interleukin-6: biology, signaling and strategies of blockade. Cytokine & growth factor reviews **26**(5), 475–487 (2015)
40. Rodrigues-Amorim, D., Rivera-Baltanas, T., Spuch, C., Caruncho, H.J., Gonzalez-Fernandez, A., Olivares, J.M., Agis-Balboa, R.C.: Cytokines dysregulation in schizophrenia: a systematic review of psychoneuroimmune relationship. Schizophrenia research **197**, 19–33 (2018)
41. Miller, B.J., Buckley, P., Seabolt, W., Mellor, A., Kirkpatrick, B.: Meta-analysis of cytokine alterations in schizophrenia: clinical status and antipsychotic effects. Biological psychiatry **70**(7), 663–671 (2011)
42. Fang, X., Zhang, Y., Fan, W., Tang, W., Zhang, C.: Interleukin-17 alteration in first-episode psychosis: a meta-analysis. Complex Psychiatry **3**(3), 135–140 (2017)
43. Çakici, N., Sutterland, A.L., Penninx, B.W., Dalm, V.A., de Haan, L., van Beveren, N.J.: Altered peripheral blood compounds in drug-naïve first-episode patients with either schizophrenia or major depressive disorder: a meta-analysis. Brain, behavior, and immunity **88**, 547–558 (2020)
44. Pandey, G.N., Ren, X., Rizavi, H.S., Zhang, H.: Proinflammatory cytokines and their membrane-bound receptors are altered in the lymphocytes of schizophrenia patients. Schizophrenia research **164**(1-3), 193–198 (2015)
45. Hope, S., Ueland, T., Steen, N.E., Dieset, I., Lorentzen, S., Berg, A.O., Agartz, I., Aukrust, P., Andreassen, O.A.: Interleukin 1 receptor antagonist and soluble tumor necrosis factor receptor 1 are associated with general severity and psychotic symptoms in schizophrenia and bipolar disorder. Schizophrenia research **145**(1-3), 36–42 (2013)
46. Pillinger, T., Osimo, E.F., Brugger, S., Mondelli, V., McCutcheon, R.A., Howes, O.D.: A meta-analysis of immune parameters, variability, and assessment of modal distribution in psychosis and test of the immune subgroup hypothesis. Schizophrenia bulletin **45**(5), 1120–1133 (2019)
47. Momtazmanesh, S., Zare-Shahabadi, A., Rezaei, N.: Cytokine alterations in schizophrenia: an updated review. Frontiers in Psychiatry **10**, 892 (2019)
48. Fond, G., d'Albis, M.-A., Jamain, S., Tamouza, R., Arango, C., Fleischhacker, W.W., Glenthøj, B., Leweke, M., Lewis, S., McGuire, P., *et al.*: The promise of biological markers for treatment response in first-episode psychosis: a systematic review. Schizophrenia bulletin **41**(3), 559–573 (2015)
49. Mondelli, V., Ciufolini, S., Belvederi Murri, M., Bonaccorso, S., Di Forti, M., Giordano, A., Marques, T.R., Zunszain, P.A., Morgan, C., Murray, R.M., *et al.*: Cortisol and inflammatory biomarkers predict poor treatment response in first episode psychosis. Schizophrenia bulletin **41**(5), 1162–1170 (2015)
50. He, X., Ma, Q., Fan, Y., Zhao, B., Wang, W., Zhu, F., Ma, X., Zhou, L.: The role of cytokines in predicting the efficacy of acute stage treatment in patients with schizophrenia. Neuropsychiatric Disease and Treatment **16**, 191 (2020)
51. Barbosa, S., Martinuzzi, E., Sommer, I.E., Dazzan, P., McGuire, P., Arango, C., Diaz-Caneja, C.M., Fleischhacker, W.W., Rujescu, D., Glenthøj, B., et al.: Baseline levels of c-reactive protein and proinflammatory cytokines are not associated with early response to amisulpride in patients with first episode psychosis: the optimise cohort study. Schizophrenia Bulletin Open (2021)
52. Pardo-de-Santayana, G., Juncal-Ruiz, M., Vázquez-Bourgon, J., Riesco-Dávila, L., de la Foz, V.O.-G., Pelayo-Terán, J.M., López-Hoyos, M., Crespo-Facorro, B.: Active psychosis and pro-inflammatory cytokines in first-episode of psychosis. Journal of Psychiatric Research **134**, 150–157 (2021)
53. Herron, J.W., Nerurkar, L., Cavanagh, J.: Neuroimmune biomarkers in mental illness. In: Biomarkers in Psychiatry, pp. 45–78. Springer, ??? (2018)
54. Martinuzzi, E., Barbosa, S., Daoudlarian, D., Ali, W.B.H., Gilet, C., Fillatre, L., Khalfallah, O., Troudet, R., Jamain, S., Fond, G., *et al.*: Stratification and prediction of remission in first-episode psychosis patients: the optimise cohort study. Translational psychiatry **9**(1), 1–13 (2019)
55. Crespo-Facorro, B., Pérez-Iglesias, R., Mata, I., Martínez-Garcia, O., Ortiz, V., Pelayo-Terán, J.M., Valdizan, E., Vazquez-Barquero, J.L.: Long-term (3-year) effectiveness of haloperidol, risperidone and olanzapine: results of a randomized, flexible-dose, open-label comparison in first-episode nonaffective psychosis. Psychopharmacology **219**(1), 225–233 (2012)
56. Stapel, B., Sieve, I., Falk, C.S., Bleich, S., Hilfiker-Kleiner, D., Kahl, K.G.: Second generation atypical antipsychotics olanzapine and aripiprazole reduce expression and secretion of inflammatory cytokines in human immune cells. Journal of psychiatric research **105**, 95–102 (2018)
57. Corsi-Zuelli, F., Loureiro, C.M., Shuhama, R., Fachim, H.A., Menezes, P.R., Louzada-Junior, P., Mondelli, V., Del-Ben, C.M.: Cytokine profile in first-episode psychosis, unaffected siblings and community-based controls: the effects of familial liability and childhood maltreatment. Psychological medicine **50**(7), 1139–1147 (2020)
58. MacDowell, K.S., García-Bueno, B., Madrigal, J.L., Parellada, M., Arango, C., Micó, J.A., Leza, J.C.: Risperidone normalizes increased inflammatory parameters and restores anti-inflammatory pathways in a model of neuroinflammation. International Journal of Neuropsychopharmacology **16**(1), 121–135 (2013)
59. Kubistova, A., Horacek, J., Novak, T.: Increased interleukin-6 and tumor necrosis factor alpha in first episode schizophrenia patients versus healthy controls. Psychiatria danubina **24**(suppl 1), 153–156 (2012)
60. Gonzalez-Vivas, C., Soldevila-Matias, P., Sparano, O., García-Martí, G., Martí-Bonmatí, L., Crespo-Facorro, B., Aleman, A., Sanjuan, J.: Longitudinal studies of functional magnetic resonance imaging in first-episode psychosis: a systematic review. European Psychiatry **59**, 60–69 (2019)
61. Kaminski, J., Mascarell-Maricic, L., Fukuda, Y., Katthagen, T., Heinz, A., Schlagenhauf, F.: Glutamate in the dorsolateral prefrontal cortex in patients with schizophrenia: A meta-analysis of 1h-magnetic resonance

spectroscopy studies. Biological Psychiatry **89**(3), 270–277 (2021)
62. Kim, M., Kim, T., Hwang, W.J., Lho, S.K., Moon, S.-Y., Lee, T.Y., Kwon, J.S.: Forecasting prognostic trajectories with mismatch negativity in early psychosis. Psychological Medicine, 1–11 (2021)
63. Schnack, H.G.: Improving individual predictions: machine learning approaches for detecting and attacking heterogeneity in schizophrenia (and other psychiatric diseases). Schizophrenia research **214**, 34–42 (2019)
64. Kwakernaak, S., van Mens, K., Cahn, W., Janssen, R., Investigators, G., *et al.*: Using machine learning to predict mental healthcare consumption in non-affective psychosis. Schizophrenia research **218**, 166–172 (2020)
65. Koutsouleris, N., Kahn, R.S., Chekroud, A.M., Leucht, S., Falkai, P., Wobrock, T., Derks, E.M., Fleischhacker, W.W., Hasan, A.: Multisite prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: a machine learning approach. The Lancet Psychiatry **3**(10), 935–946 (2016)
66. Koutsouleris, N., Kambeitz-Ilankovic, L., Ruhrmann, S., Rosen, M., Ruef, A., Dwyer, D.B., Paolini, M., Chisholm, K., Kambeitz, J., Haidl, T., *et al.*: Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: a multimodal, multisite machine learning analysis. JAMA psychiatry **75**(11), 1156–1172 (2018)
67. Darcy, A.M., Louie, A.K., Roberts, L.W.: Machine learning and the profession of medicine. Jama **315**(6), 551–552 (2016)
68. Obermeyer, Z., Emanuel, E.J.: Predicting the future—big data, machine learning, and clinical medicine. The New England journal of medicine **375**(13), 1216 (2016)
69. Amato, F., López, A., Peña-Méndez, E.M., Vaňhara, P., Hampl, A., Havel, J.: Artificial neural networks in medical diagnosis. Elsevier (2013)
70. Leighton, S.P., Upthegrove, R., Krishnadas, R., Benros, M.E., Broome, M.R., Gkoutos, G.V., Liddle, P.F., Singh, S.P., Everard, L., Jones, P.B., *et al.*: Development and validation of multivariable prediction models of remission, recovery, and quality of life outcomes in people with first episode psychosis: a machine learning approach. The Lancet Digital Health **1**(6), 261–270 (2019)
71. Lanillos, P., Oliva, D., Philippsen, A., Yamashita, Y., Nagai, Y., Cheng, G.: A review on neural network models of schizophrenia and autism spectrum disorder. Neural Networks **122**, 338–363 (2020)
72. Koppe, G., Meyer-Lindenberg, A., Durstewitz, D.: Deep learning for small and big data in psychiatry. Neuropsychopharmacology **46**(1), 176–190 (2021)
73. Vieira, S., Pinaya, W.H., Mechelli, A.: Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. Neuroscience & Biobehavioral Reviews **74**, 58–75 (2017)
74. Cortes-Briones, J.A., Tapia-Rivas, N.I., D'Souza, D.C., Estevez, P.A.: Going deep into schizophrenia with artificial intelligence. Schizophrenia Research (2021)
75. de Nijs, J., Burger, T.J., Janssen, R.J., Kia, S.M., van Opstal, D.P., de Koning, M.B., de Haan, L., Cahn, W., Schnack, H.G.: Individualized prediction of three-and six-year outcomes of psychosis in a longitudinal multicenter study: a machine learning approach. npj Schizophrenia **7**(1), 1–11 (2021)
76. Stojanovic, A., Martorell, L., Montalvo, I., Ortega, L., Monseny, R., Vilella, E., Labad, J.: Increased serum interleukin-6 levels in early stages of psychosis: associations with at-risk mental states and the severity of psychotic symptoms. Psychoneuroendocrinology **41**, 23–32 (2014)
77. Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., *et al.*: The mini-international neuropsychiatric interview (mini): the development and validation of a structured diagnostic psychiatric interview for dsm-iv and icd-10. Journal of clinical psychiatry **59**(20), 22–33 (1998)
78. Andreasen, N.C., Carpenter Jr, W.T., Kane, J.M., Lasser, R.A., Marder, S.R., Weinberger, D.R.: Remission in schizophrenia: proposed criteria and rationale for consensus. American Journal of Psychiatry **162**(3), 441–449 (2005)
79. Kemmler, G., Hummer, M., Widschwendter, C., Fleischhacker, W.W.: Dropout rates in placebo-controlled and active-control clinical trials of antipsychotic drugs: a meta-analysis. Archives of general psychiatry **62**(12), 1305–1312 (2005)
80. Little, R.J., D'Agostino, R., Cohen, M.L., Dickersin, K., Emerson, S.S., Farrar, J.T., Frangakis, C., Hogan, J.W., Molenberghs, G., Murphy, S.A., *et al.*: The prevention and treatment of missing data in clinical trials. New England Journal of Medicine **367**(14), 1355–1360 (2012)
81. Kwak, S.K., Kim, J.H.: Statistical data preparation: management of missing values and outliers. Korean journal of anesthesiology **70**(4), 407 (2017)
82. Rabinowitz, J., Davidov, O.: The association of dropout and outcome in trials of antipsychotic medication and its implications for dealing with missing data. Schizophrenia bulletin **34**(2), 286–291 (2008)
83. Leurent, B., Gomes, M., Faria, R., Morris, S., Grieve, R., Carpenter, J.R.: Sensitivity analysis for not-at-random missing data in trial-based cost-effectiveness analysis: a tutorial. Pharmacoeconomics **36**(8), 889–901 (2018)
84. Kia, S.M.: PROMISSING: Pruning Missing Values in Neural Networks. submitted (2021)
85. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
86. Agarap, A.F.: Deep learning using rectified linear units (relu). arXiv preprint arXiv:1803.08375 (2018)
87. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
88. Mothilal, R.K., Sharma, A., Tan, C.: Explaining machine learning classifiers through diverse counterfactual explanations. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, pp. 607–617 (2020)
89. Ojala, M., Garriga, G.C.: Permutation tests for studying classifier performance. Journal of Machine Learning Research **11**(6) (2010)
90. Nichols, T.E., Holmes, A.P.: Nonparametric permutation tests for functional neuroimaging: a primer with

examples. Human brain mapping **15**(1), 1–25 (2002)
91. Justice, A.C., Covinsky, K.E., Berlin, J.A.: Assessing the generalizability of prognostic information. Annals of internal medicine **130**(6), 515–524 (1999)
92. Schnack, H.G., Kahn, R.S.: Detecting neuroimaging biomarkers for psychiatric disorders: sample size matters. Frontiers in psychiatry **7**, 50 (2016)
93. Çakici, N., Van Beveren, N., Judge-Hundal, G., Koola, M., Sommer, I.: An update on the efficacy of anti-inflammatory agents for patients with schizophrenia: a meta-analysis. Psychological medicine **49**(14), 2307–2319 (2019)
94. Jeppesen, R., Christensen, R.H., Pedersen, E.M., Nordentoft, M., Hjorthøj, C., Köhler-Forsberg, O., Benros, M.E.: Efficacy and safety of anti-inflammatory agents in treatment of psychotic disorders–a comprehensive systematic review and meta-analysis. Brain, behavior, and immunity **90**, 364–380 (2020)
95. Goorbergh, R.v.d., van Smeden, M., Timmerman, D., Van Calster, B.: The harm of class imbalance corrections for risk prediction models: illustration and simulation using logistic regression. arXiv preprint arXiv:2202.09101 (2022)
96. Smith, R.C., Leucht, S., Davis, J.M.: Maximizing response to first-line antipsychotics in schizophrenia: a review focused on finding from meta-analysis. Psychopharmacology **236**(2), 545–559 (2019)
97. Woo, C.-W., Chang, L.J., Lindquist, M.A., Wager, T.D.: Building better biomarkers: brain models in translational neuroimaging. Nature neuroscience **20**(3), 365–377 (2017)
98. Kay, S.R., Fiszbein, A., Opler, L.A.: The positive and negative syndrome scale (panss) for schizophrenia. Schizophrenia bulletin **13**(2), 261–276 (1987)
99. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Software available from tensorflow.org (2015). https://www.tensorflow.org/
100. Chollet, F., et al.: Keras. GitHub (2015). https://github.com/fchollet/keras
101. Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., del Río, J.F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E.: Array programming with NumPy. Nature **585**(7825), 357–362 (2020). doi:10.1038/s41586-020-2649-2
102. McKinney, W., *et al.*: Data structures for statistical computing in python. In: Proceedings of the 9th Python in Science Conference, vol. 445, pp. 51–56 (2010). Austin, TX

Appendix
A.1

Table 5 Total PANSS Scores for Patients With and Without Missing Values

		Total PANSS	
Modality	**% missing**	**Patients without missing values**	**Patients with missing values**
Demographics	0%	77.9 (18.7)	-
Diagnosis	4%	77.7 (18.7)	82.9 (19)
Lifestyle	88%	80.8 (16.9)	77.5 (18.9)
Somatic	21%	78.0 (19.1)	77.6 (17.1)
Treatments	2%	78.0 (18.7)	83.4 (21.7)
MINI	4%	77.1 (18.3)	98.2 (17.7)
Cytokines	21%	79.2 (19.0)	74.9 (18.6)
PSP	3%	77.8 (18.7)	80.4 (17.4)
CGI	1%	78.0 (18.7)	68.8 (19.9)
CDSS	2%	77.7 (18.8)	87.3 (12.1)
SWN	9%	77.1 (18.7)	86.8 (15.9)

The data are presented as mean (SD) and based on all 454 patients that had PANSS scores measured at the start of the OPTiMiSE study. patients that had at least one missing value for any given modality were included in the *patients with missing values* category. Patients with complete data were included in the *patients without missing values* category. The *% missing* is thus the proportion of patients that had at least one missing value for that modality. The *PANSS (total)* is the sum of 30 items included in the Positive and Negative Syndrome Scale [98]. As the scale ranges from 1-7, theoretical scores range from 30 to 210.

A.2

The deep neural network was implemented using the packages `tensorflow-cpu 2.4.1` [99], `keras 2.4.3` [100], `numpy 1.19.5` [101], `pandas 1.0.5` [102] and `scikit-learn 0.23.1` [85] using Python 3.8 as the interpreter. Data analyses were performed on a Windows X64-based machine with an Intel i7-6700 CPU and 16 GB RAM.

A.3

Table 6 Classification Performance at Week 4 for Two Different Strategies to Handle Missing Values. AUC scores are shown as mean (statistical deviation, SD). AUC scores are shown for two different strategies to handle missing values: k-NN impute and neutralization. Sensitivity is also called *recall* or *true positive rate (TPR)*, while specificity is also called *true negative rate* (TNR) or *selectivity*.

	Imputation			**Neutralization**		
Modality	**AUC**	**Sensitivity**	**Specificity**	**AUC**	**Sensitivity**	**Specificity**
Demographics	0.60 (0.02)	0.64	0.47	0.60 (0.02)	0.65	0.46
Diagnosis	0.58 (0.02)	0.66	0.44	0.58 (0.03)	0.66	0.46
Lifestyle	0.47 (0.03)	0.55	0.41	0.51 (0.03)	0.53	0.49
Somatic	0.49 (0.03)	0.61	0.37	0.50 (0.03)	0.60	0.40
Treatments	0.51 (0.02)	0.51	0.52	0.52 (0.02)	0.53	0.53
MINI	0.56 (0.01)	0.68	0.43	0.57 (0.02)	0.70	0.44
Cytokines	0.58 (0.01)	0.71	0.34	0.58 (0.02)	0.71	0.33
PANSS	0.60 (0.02)	0.72	0.42	0.60 (0.02)	0.70	0.39
PSP	0.58 (0.02)	0.56	0.57	0.59 (0.02)	0.56	0.58
CGI	0.55 (0.02)	0.51	0.65	0.55 (0.01)	0.50	0.55
CDSS	0.49 (0.03)	0.65	0.37	0.50 (0.02)	0.65	0.40
SWN	0.50 (0.03)	0.65	0.35	0.52 (0.02)	0.64	0.39
All	0.65 (0.02)	0.86	0.33	0.66 (0.01)	0.85	0.32

A.4

Table 7 Classification Performance at Week 10 for Two Different Strategies to Handle Missing Values. AUC scores are shown as mean (statistical deviation, SD).

	Imputation			**Neutralization**		
Modality	**AUC**	**Sensitivity**	**Specificity**	**AUC**	**Sensitivity**	**Specificity**
Demographics	0.54 (0.05)	0.57	0.54	0.50 (0.04)	0.48	.50
Diagnosis	0.49 (0.07)	0.48	0.50	0.59 (0.01)	0.88	0.31
Lifestyle	0.55 (0.06)	0.56	0.53	0.54 (0.07)	0.55	0.55
Somatic	0.41 (0.05)	0.42	0.48	0.42 (0.08)	0.45	0.43
Treatments	0.48 (0.08)	0.53	0.42	0.52 (0.07)	0.54	0.55
MINI	0.43 (0.06)	0.44	0.43	0.41 (0.06)	0.44	0.38
Cytokines	0.66 (0.03)	0.60	0.61	0.67 (0.04)	0.61	0.61
PANSS	0.51 (0.06)	0.48	0.52	0.58 (0.06)	0.55	0.56
PSP	0.52 (0.06)	0.53	0.48	0.51 (0.06)	0.51	0.52
CGI	0.49 (0.09)	0.51	0.49	0.48 (0.09)	0.49	0.49
CDSS	0.62 (0.04)	0.67	0.57	0.60 (0.05)	0.65	0.52
SWN	0.60 (0.04)	0.53	0.62	0.59 (0.05)	0.51	0.61
All	0.63 (0.06)	0.54	0.69	0.63 (0.07)	0.54	0.64

A.5

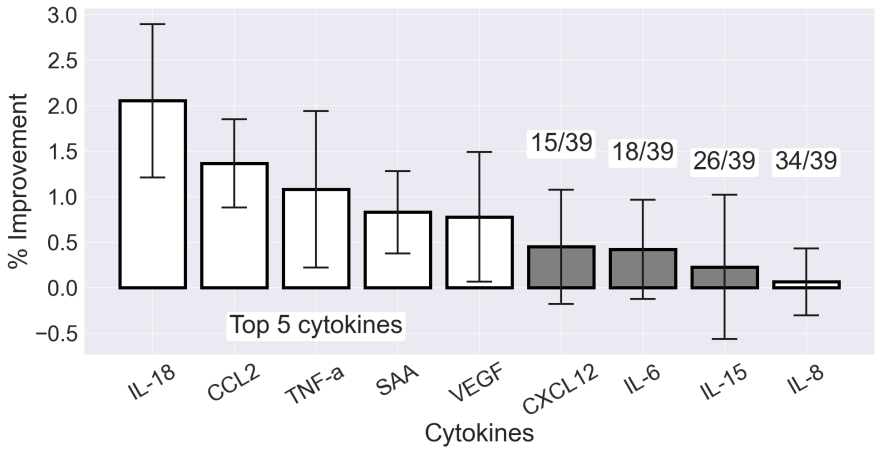

Figure 4 Individual Cytokines' Contribution to Prediction Certainty. Improvement in prediction certainty (%) for the top 5 performing cytokines (white) and cytokines IL-6, IL-8, IL-15 and CXCL12 (gray). The rank (out of 39 cytokines) is indicated by the white box.

A.6

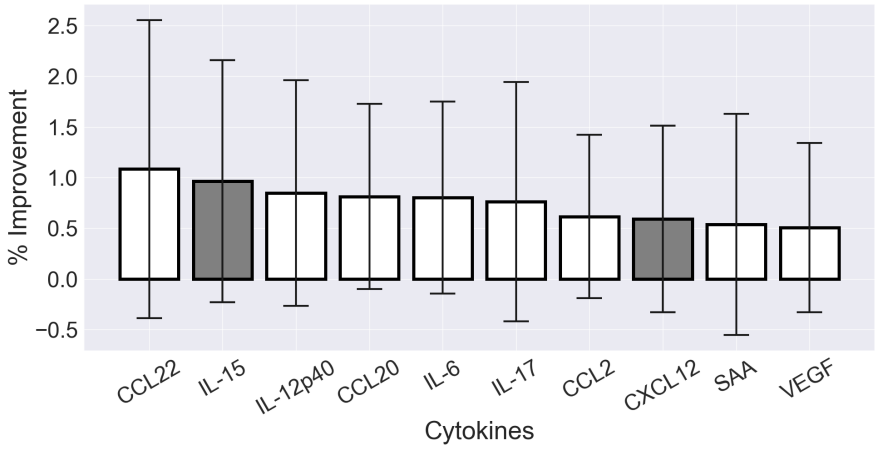

Figure 5 Individual Cytokines' Contribution to Prediction Certainty for a Cluster of Patients. Improvement in prediction certainty (%) for the top 10 performing cytokines (white) and cytokines IL-6 and CXCL12 (gray) implicated in psychosis prognosis prediction for the a cluster of 97 FEP patients.