The classification model of social functional remission in patients with schizophrenia treated with paliperidone palmitate using machine learning technique

Hisanori Kobayashi, Takashi Ohnishi, Ryoko Nakagawa, Kazutake Yoshizawa

Evidence Generation Dept., Medical Affairs Div., Janssen Pharmaceutical K.K.

Background

- The improvement of social function is an important outcome for patients with schizophrenia.
- Our previous naturalistic observational study has suggested that early improvement of positive symptoms should be related to the functional remission. [1]
- However, several important factors, such as socio-demographic and supports from caregiver, are possibly associated with functional remission and also need to be evaluated.

To date numerous methods for classification problem are available and applied to various research areas.

- Studies suggested the usefulness of application of machine learning classification techniques to psychiatric data for obtaining a higher flexibility and precision.

Objectives

- To evaluate socio-demographic data and neuropsychological measurement scores by developing a classification model of social functional remission in patients with schizophrenia by using machine learning approach.

Methods

Study design

- This study is a secondary (subgroup) data analysis utilizing previously conducted, randomized, double-blind, parallel-group, multicenter, clinical trial (PSY-3011: EudraCT registration number: 2011-004889-15). [2] The original trial has been conducted to evaluate the non-inferiority of 3-month paliperidone palmitate formulation compared with the 1-month formulation.

Participants

- Diagnosed with schizophrenia
- Received paliperidone palmitate
- Scored ≤70 in personal and social performance scale (PSP) at baseline
- Participants who reported data of baseline and 65 weeks (see variable section)

Outcome variables

- Reference (response) variable
  - Social function remission [Y/N] (Social function remission was defined as participants who scored PSP ≤70 at the 65-week end point)

- Attributes (explanatory) variables
  - Baseline demographics
    - Age
    - Sex
    - Races (White [reference], Asian, Black, Other)
    - Type of schizophrenia (Catatonic, Disorganize, Paranoid [reference], Residual, Undifferentiated)
    - Age at first diagnosed
    - Family psychiatric history: Y/N
  - Baseline Neuropsychiatric Measures
    - Baseline PSP (subscale; Social useful activity, Social relationship, Self care, Aggressive behavior)
    - Early improvement of positive and negative symptoms (%-change between open label and double-blind baseline in sub-scores of positive and negative syndrome scale: PANSS general, PANSS positive, PANSS negative)
  - Caregiver involvement
    - Sub-scores of involvement evaluation questionnaire: IEQ-Supervision, IEQ-Tension, IEQ-Urging, IEQ-Worrying

Statistical analysis

A set of logistic regressions is first performed to examine whether the addition of measurements is worthwhile for the better prediction of social function remission.

- Patients’ demographics, neuropsychological measure, and IEQ were sequentially added to the model.
  - Model 1: Social function remission = BL demographics
  - Model 2: Social function remission = BL demographics + neuropsychological measure + IEQ
  - Model 3: Social function remission = BL demographics + neuropsychological measure + IEQ

Then, we applied a Support Vector Machine (SVM) algorithm to re-evaluate the model to examine whether the model predictive performance was different between the algorithms.

- The SVM training was performed with Gaussian radial basis kernel function and 10-fold cross validation.
- Finally, we applied feature subset selection methodology to the model to identify importance of each attribute and within the data
- Logistic regressions and SVM were performed using R and Matlib.

Results

Analytical sample (Figure 1)

- Among 1429 samples about 30% (n=428) of subject were non-social function remission at baseline and reported all information for model building.
- Data from 428 participants, including 100 who achieved remission at 65-week, were utilized for the analysis.

Logistic model

- The predictive performance of the first logistic regression model, including only demographic variables, was not adequate (area under curve; AUC = 0.606).
- After neuropsychiatric measures were included to the model, the performance was improved (AUC = 0.737).
- Addition of IEQ measures to the second model also contributed to better performance (AUC = 0.755). Younger age (p = 0.017), Schizophrenia type (Residual vs. Paranoid, p = 0.043) higher social activity (p = 0.003) and social relationship (p = 0.003), and higher caregiver support (p = 0.007) were significantly associated with higher social function remission.

SVM model and feature subset selection

- The third logistic model was tested using the SVM and the results showed the model classifying the social function remission adequately (C = 5, sigma = 0.050, accuracies of training and validation set were 0.916 and 0.743, respectively).
- Finally, feature subset selection using learning vector quantization model was applied for examine the contribution of each variable.
- The top 5 most important variables associated with social function remission were high social relationship, improvement of negative symptom, younger age, low aggressive behavior, and high support of supervision. (Figure 2)

Discussion/Conclusion

- Both logistic regression and SVM approaches showed relatively similar results and supported the usability of the SVM algorithm for the classification of social function remission.
- Analyses supported that the addition of neuropsychiatric and caregiving information improves the model performance, and suggested that these information should be important factors for explaining future social functional remission.

- While the results revealed some significant factors associated with functional remission, it also suggested room for improvement in the classification performance.

- The 2 approaches suggested different degrees of contributions through the variables in the model. This may suggests that conducting multiple approaches and interpreted results in multiple ways should be important during building the statistical model.

References/COI


Personal conflicts of interest / Funding disclosure

All authors are full time employees at Janssen Pharmaceutical K.K. The company sponsored this study.

References