## Principal Investigator

**First Name:** Raviv
**Last Name:** Pryluk
**Degree:** PhD
**Primary Affiliation:** PhaseV Trials
**E-mail:** raviv@phasevtrials.com
**State or Province:** MA
**Country:** United States

## General Information

**Key Personnel (other than PI):**
**First Name:** Elad
**Last name:** Berkman
**Degree:** MSc
**Primary Affiliation:** PhaseV Trials
**SCOPUS ID:**
**Requires Data Access?** Yes

**First Name:** Oshri
**Last name:** Machluf
**Degree:** BA
**Primary Affiliation:** PhaseV Trials
**SCOPUS ID:**
**Requires Data Access?** Yes

**First Name:** Gal
**Last name:** Shoham
**Degree:** BA
**Primary Affiliation:** PhaseV Trials
**SCOPUS ID:**
**Requires Data Access?** Yes

**Are external grants or funds being used to support this research?:** No external grants or funds are being used to support this research.
**How did you learn about the YODA Project?:** PubMed

## Conflict of Interest

https://yoda.yale.edu/wp-content/uploads/2024/02/YODA_TIB_COI.pdf
https://yoda.yale.edu/wp-content/uploads/2024/03/YODA_RP_COI.pdf
https://yoda.yale.edu/wp-content/uploads/2024/03/YODA_EB_COI.pdf
https://yoda.yale.edu/wp-content/uploads/2024/03/YODA_OM_COI.pdf
https://yoda.yale.edu/wp-content/uploads/2024/03/YODA_GS_COI.pdf

## Certification

**Certification:** All information is complete; I (PI) am responsible for the research; data will not be used to support litigious/commercial aims.
**Data Use Agreement Training:** As the Principal Investigator of this study, I certify that I have

completed the YODA Project Data Use Agreement Training

1. NCT00518323 - R076477PSZ3001 - A Randomized, Multicenter, Double-Blind, Weight-Based, Fixed-Dose, Parallel-Group, Placebo-Controlled Study of the Efficacy and Safety of Extended Release Paliperidone for the Treatment of Schizophrenia in Adolescent Subjects, 12 to 17 Years of Age
2. NCT00334126 - R076477SCH3015 - A Randomized, Double-blind, Placebo-controlled, Parallel Group Study to Evaluate the Efficacy and Safety of Paliperidone ER Compared to Quetiapine in Subjects With an Acute Exacerbation of Schizophrenia
3. NCT00085748 - R076477-SCH-302 - A Randomized, 6-Week Double-Blind, Placebo-Controlled Study With an Optional 24-Week Open-Label Extension to Evaluate the Safety and Tolerability of Flexible Doses of Paliperidone Extended Release in the Treatment of Geriatric Patients With Schizophrenia
4. NCT00752427 - R076477-SCH-702 - 24 week extension of NCT00085748: A Randomized, 6-Week Double-Blind, Placebo-Controlled Study With an Optional 24-Week Open-Label Extension to Evaluate the Safety and Tolerability of Flexible Doses of Paliperidone Extended Release in the Treatment of Geriatric Patients With Schizophrenia
5. NCT00078039 - R076477-SCH-303 - Trial Evaluating Three Fixed Dosages of Paliperidone Extended-Release (ER) Tablets and Olanzapine in the Treatment of Patients With Schizophrenia
6. NCT00083668 - R076477-SCH-305 - A Randomized, Double-blind, Placebo- and Active-controlled, Parallel-group, Dose-response Study to Evaluate the Efficacy and Safety of 3 Fixed Dosages of Paliperidone Extended Release (ER) Tablets and Olanzapine, With Open-label Extension, in the Treatment of Patients With Schizophrenia
7. NCT00488319 - R076477PSZ3002 - A 2-Year, Open-Label, Single-Arm Safety Study of Flexibly Dosed Paliperidone Extended Release (1.5-12 mg/day) in the Treatment of Adolescents (12 to 17 Years of Age) With Schizophrenia
8. NCT01009047 - R076477PSZ3003 - A Randomized, Multicenter, Double-Blind, Active-Controlled, Flexible-Dose, Parallel-Group Study of the Efficacy and Safety of Prolonged Release Paliperidone for the Treatment of Symptoms of Schizophrenia in Adolescent Subjects, 12 to 17 Years of Age
9. NCT03345342 - R092670PSY3015 - A Double-blind, Randomized, Active-controlled, Parallel-group Study of Paliperidone Palmitate 6-Month Formulation
10. NCT02713282 - R092670SCH3015 - A 52-Week, Open-Label, Prospective, Multicenter, International Study of a Transition to the Paliperidone Palmitate 3-Month Formulation In Patients With Schizophrenia Previously Stabilized on the Paliperidone Palmitate 1-Month Formulation
11. NCT01515423 - R092670PSY3011 - A Randomized, Multicenter, Double-Blind, Non-inferiority Study of Paliperidone Palmitate 3 Month and 1 Month Formulations for the Treatment of Subjects With Schizophrenia

**What type of data are you looking for?:** Individual Participant-Level Data, which includes Full CSR and all supporting documentation

## Research Proposal

## Project Title

Use of machine learning for discerning prognostic and predictive features in schizophrenia disease and evaluating their generalizability

**Narrative Summary:**

Schizophrenia is a highly heterogenous disorder. In this retrospective analysis of clinical trial data we aim to discern factors that influence Schizophrenia disease progression, either regardless of or

considering treatment, by applying advanced machine learning methods to clinical trial data.
There have been made attempts at constructing generalizable clinical prediction models for different diseases. To our knowledge, no such model has been successful constructed for Schizophrenia Disease.
Discerning prognostic and predictive features may contribute to an improved understanding of Schizophrenia Disease heterogeneity. Such a model may be integrated within clinical development processes and considered in strategic trial decision making, and potentially aid in trial design and improve selection of trial populations. This could potentially have a positive impact on Schizophrenia drug development.

**Scientific Abstract:**

Background: Schizophrenia is a highly heterogenous disorder for which, to our knowledge, no generalizable clinical prediction model is available. Building on the work of Chekroud et al, we hypothesize that by revisiting the machine learning models constructed and developing new models, new findings may be produced such that may enable the identification of discerning prognostic and predictive features that may be generalizable when implemented across several clinical trial data sets.
Objective: The research objectives of this project are to discern prognostic and predictive features in schizophrenia disease using machine learning and evaluate their generalizability by implementing them across different clinical trial data sets.
Study Design: Trial data will be preprocessed according to inclusion and exclusion criteria and important missing baseline values will be imputed. Then, the processed data will be analyzed descriptively and the control population from each trial will be employed in the construction of a prognostic models and their performance will be compared to each other. Next, causal machine learning will be applied to estimate the conditional average treatment effect (CATE). CATE estimators will be applied between trials to assess the quality of the prediction.
Participants; This study will include subjects from the five trials analyzed by Chekroud et al and an additional 7 studies of the same drug. The analyzed cohort will include (i) Subjects with a current DSM-IV diagnosis of schizophrenia (and type) at the start of the respective trial; (ii) subjects were randomized to an antipsychotic medication or placebo; (iii) subjects for whom there exists the 4-week calculated PANSS score. Those subjects for whom insufficient data exists will be excluded.
Primary outcome measure: Remission in Schizophrenia Working Group criteria (RSWG)- early treatment response at 4 weeks follow up time point
Secondary Outcome Measures: (i) 25% symptom reduction (binary); (ii) 50% symptom reduction (binary); (iii) Baseline-adjusted percent change in symptoms (continuous).
Statistical Analysis: The control patients from each trial will be used to build a prognostic model for the relevant endpoints, using random-forests, BART (Bayesian additive regression trees), AGLM (accurate GLM, a method using LASSO on bin-encoded features to obtain a piecewise-constant estimator) etc. The performance of the models trained on each of the trials will be tested on the remaining trials. Performance will be compared using AUC for binary EPs and RMSE for continuous EPs, we will test the significance of these comparisons using a permutation test (which will randomize the predictions across patients to generate the null hypothesis of independence between prediction and outcome). We will also apply causal-ML methods to estimate the CATE (conditional average treatment effect), including Causal-Forest and meta-learners. The CATE estimators from each trial will be applied to other trials, and the quality of their prediction will be assessed using the curve of cumulated ATE when sorted by the predicted CATE. The area between this curve and the constant ATE line (ABC for area-between-curves) is a measure for the explained heterogeneity in treatment effect. This will be compared to the ABC within trial (on held-out validation data). Results will be tested using a permutation test as well.

**Brief Project Background and Statement of Project Significance:**

By discerning prognostic and predictive features that are generalizable across different clinical trials, we may contribute to an improved understanding of Schizophrenia Disease heterogeneity. Models constructed based on such features may be integrated within clinical development processes and

considered in strategic trial decision making, hence potentially aid in trial design and improve selection of trial populations. This can have a positive impact on Schizophrenia drug development.

**Specific Aims of the Project:**

The research objectives of this project are to discern prognostic and predictive features in schizophrenia disease using machine learning and evaluate their generalizability by implementing them across different clinical trial data sets.

**Study Design:**

Methodological research

**What is the purpose of the analysis being proposed? Please select all that apply.**

Confirm or validate previously conducted research on treatment effectiveness

Develop or refine statistical methods

Research on clinical prediction or risk prediction

## Research Methods

**Data Source and Inclusion/Exclusion Criteria to be used to define the patient sample for your study:**

Inclusion criteria:
1. Subjects with a current DSM-IV diagnosis of schizophrenia (and type) at the start of the respective trial
2. Subjects were randomized to an antipsychotic medication or placebo;
3. Subjects for whom there exists the 4-week calculated PANSS score

Exclusion criteria:
1. Subjects for whom insufficient data exists

**Primary and Secondary Outcome Measure(s) and how they will be categorized/defined for your study:**

Primary outcome measure:
Remission in Schizophrenia Working Group criteria (RSWG)- early treatment response at 4 weeks follow up time point

Secondary outcome measures:
• 25% symptom reduction (binary)
• 50% symptom reduction (binary)
• Baseline-adjusted percent change in symptoms (continuous)

**Main Predictor/Independent Variable and how it will be categorized/defined for your study:**

Main predictor defined according to the Remission in Schizophrenia Working Group (RSWG) criteria based on 7 PANSS criteria ratings and duration:
1. Delusions
2. Hallucinations
3. Bizarre behavior
4. Poor attention

5. Blunted affect
6. Alogia
7. Anhedonia

**Other Variables of Interest that will be used in your analysis and how they will be categorized/defined for your study:**

1. Demographic: Sex, Race, Age, Age at diagnosis, number of hospitalizations
2. Medications: drug, dose
3. Vital signs: Weight, Height, BMI, Pulse, systolic and diastolic blood pressure
4. Blood chemistry: Alkaline phosphatase, Aminotransferase, Aspartate aminotransferase, Bilirubin Calcium, Chloride, Creatinine, Gamma glutamyl transferase, Potassium, Protein, Sodium
5. Complete blood count with differential: Basophils, Eosinophils, Hematocrit, Hemoglobin, Lymphocytes, Monocytes, Neutrophils, Platelets, Red blood cells, White blood cells
6. Urinalysis: Epithelial cells, leukocyte esterase, specific gravity, bacteria, bilirubin, crystals, glucose, ketones, nitrite, Ph, protein, drug (yes/no)
7. Clinical global impression score (CGI)
8. Barnes Akathisia Rating Scale: Objective, Awareness of restlessness, distress related to restlessness, Global clinical rating of Akathisia
9. Abnormal Involuntary Movement Scale: Facial expression, Lips, Jaw, Tongue, Upper arm, Lower body, Neck shoulders hips, Severity of abnormal movements, Incapacitation due to abnormal movements, Awareness of abnormal movements, Problems with teeth, Wears dentures, AIMS total
10. Simpson-Angus Scale: Gait, Arm dropping, Shoulder shaking, Elbow rigidity, Wrist rigidity, Leg pendulousness, Head dropping, Glabella tap, Tremor, Salivation, SAS total
11. Positive and Negative Syndrome Scale (PANSS):
11a. PANSS Total
11b. General: Somatic concerns, Anxiety, Guilt feelings, Tension, Mannerisms and posturing, Depression, Motor retardation, Uncooperativeness, Unusual thought content, Disorientation, Poor attention, Lack of judgement and insight, Disturbance of volition, Poor impulse control, Preoccupation, Active social avoidance, PANSS general total
11c. Positive symptoms: Delusions, Conceptual disorganization, Hallucinatory behavior, Excitement, Grandiosity, Suspiciousness/persecution, Hostility, PANSS positive total
11d. Negative symptoms: Blunted affect, Emotional withdrawal, Poor rapport, Passive/Apathetic social withdrawal, Difficulty in abstract thinking, Lack of spontaneity/ flow in conversation, Stereotyped thinking, PANSS negative Total

**Statistical Analysis Plan:**

The analysis will include preprocessing of the trials' data, including imputation of important missing baseline values using either mean/median imputation, MICE and/or MI.
The processed data will first be analyzed descriptively - comparing the trials' populations and marginal associations between baseline features and endpoints (EPs).
We will then use the control patients in each trial to build a prognostic model for the relevant EPs, using random-forests, BART (Bayesian additive regression trees), AGLM (accurate GLM, a method using LASSO on bin-encoded features to obtain a piecewise-constant estimator) etc. The performance of the models trained on each of the trials will be tested on the remaining trials. Performance will be compared using AUC for binary EPs and RMSE for continuous EPs, we will test the significance of these comparisons using a permutation test (which will randomize the predictions across patients to generate the null hypothesis of independence between prediction and outcome).
Additionally, we will apply causal-ML methods to estimate the CATE (conditional average treatment effect), including Causal-Forest and meta-learners. the CATE estimators from each trial will be applied to other trials, and the quality of their prediction will be assessed using the curve of cumulated ATE when sorted by the predicted CATE. the area between this curve and the constant ATE line (ABC for area-between-curves) is a measure for the explained heterogeneity in treatment effect. This will be compared to the ABC within trial (on held-out validation data). Results will be tested using a permutation test as well.
We intend to apply this analysis both within trials involving similar populations (namely age groups), and across types of sub-populations.

**Software Used:**

Python

**Project Timeline:**

Anticipated project start date: May 1, 2024
Analysis completion date: November 1, 2024
Manuscript draft date: January 1, 2025
First submission for publication date: March 1, 2025
Results reported to YODA date: April 30, 2025

**Dissemination Plan:**

We anticipate a minimum of one manuscript that would target Psychiatrists AND/OR Biostatisticians and the Cherkoud et al team we referenced in the project proposal.
Suitable journals may be Science, PNAS (Proceedings of the National Academy of Sciences), Statistics in Medicine.