## Principal Investigator

**First Name:** Ives
**Last Name:** Cavalcante Passos
**Degree:** MD, PhD
**Primary Affiliation:** Federal University of Rio Grande do Sul
**E-mail:** zimermanaline@gmail.com
**State or Province:** Rio Grande do Sul
**Country:** Brazil

## General Information

**Key Personnel (other than PI):**
**First Name:** Aline
**Last name:** Zimerman
**Degree:** BSc
**Primary Affiliation:** Federal University of Rio Grande do Sul
**SCOPUS ID:** 57203139232
**Requires Data Access?** Yes

**First Name:** Bruno
**Last name:** Montezano
**Degree:** BSc, MSc
**Primary Affiliation:** Federal University of Rio Grande do Sul
**SCOPUS ID:** 58609170400
**Requires Data Access?** Yes

**First Name:** Pedro
**Last name:** Lemos Ballester
**Degree:** PhD
**Primary Affiliation:** The Hospital for Sick Children
**SCOPUS ID:**
**Requires Data Access?** Yes

**First Name:** Thiago Henrique
**Last name:** Roza
**Degree:** MD, PhD
**Primary Affiliation:** Universidade Federal do Paraná
**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?:** Scientific Publication

## Conflict of Interest

https://yoda.yale.edu/wp-content/uploads/2024/02/SV_57KskaKADT3U9Aq-R_6Lq8IgIVcJB0RqN_240219_121337.pdf
https://yoda.yale.edu/wp-content/uploads/2024/02/SV_57KskaKADT3U9Aq-R_52DMcqWTAHzjqPV-1.pdf
https://yoda.yale.edu/wp-content/uploads/2024/02/SV_57KskaKADT3U9Aq-R_6rptOW30vEeXLHV.pdf

https://yoda.yale.edu/wp-content/uploads/2024/02/SV_57KskaKADT3U9Aq-R_59aLwSmqZwJ5wuo.pdf
https://yoda.yale.edu/wp-content/uploads/2024/02/coi_yoda_BM.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 - 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 - 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 - 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. NCT00650793 - A Randomized, DB, PC and AC, Parallel Group, Dose-Response Study to Evaluate the Efficacy and Safety of 3 Fixed Dosages of Extended Release OROS Paliperidone (6, 9, 12 mg/Day) and Olanzapine (10 mg/Day), With Open-Label Extension, in the Treatment of Subjects With Schizophrenia - Open Label Phase
5. NCT00083668 - 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
6. NCT00391222 - A Randomized, Double Blind, Placebo and Active Controlled Parallel Group Study to Evaluate the Efficacy and Safety of Risperidone Long-acting Injectable (LAI) for the Prevention of Mood Episodes in the Treatment of Subjects With Bipolar I Disorder
7. NCT00076115 - Research on the Effectiveness of Risperidone in Bipolar Disorder in Adolescents and Children (REACH): A Double-Blind, Randomized, Placebo-Controlled Study of the Efficacy and Safety of Risperidone for the Treatment of Acute Mania in Bipolar I Disorder
8. NCT00132678 - A Randomized, Double-blind, Placebo-controlled Study to Explore the Efficacy and Safety of Risperidone Long-acting Intramuscular Injectable in the Prevention of Mood Episodes in Bipolar 1 Disorder, With Open-label Extension
9. NCT00094926 - A Prospective, Randomized, Double-blind, Placebo-controlled Study of the Effectiveness and Safety of RISPERDAL CONSTA Augmentation in Adult Patients With Frequently-relapsing Bipolar Disorder
10. NCT00253162 - The Efficacy And Safety Of Flexible Dose Ranges Of Risperidone Versus Placebo Or Haloperidol In The Treatment Of Manic Episodes Associated With Bipolar I Disorder
11. NCT00249236 - The Efficacy And Safety Of Flexible Dosage Ranges Of Risperidone Versus Placebo In The Treatment Of Manic Or Mixed Episodes Associated With Bipolar I Disorder
12. NCT00250367 - The Safety And Efficacy Of Risperdal (Risperidone) Versus Placebo As Add-On Therapy To Mood Stabilizers In The Treatment Of The Manic Phase Of Bipolar Disorder
13. NCT00253149 - The Safety And Efficacy Of Risperdal (Risperidone) Versus Placebo Versus Haloperidol As Add-On Therapy To Mood Stabilizers In The Treatment Of The Manic Phase Of Bipolar Disorder
14. NCT00257075 - The Efficacy And Safety Of Flexible Dosage Ranges Of Risperidone Versus Placebo In The Treatment Of Manic Episodes Associated With Bipolar I Disorder
15. The efficacy and safety of flexible dose ranges of risperidone vs. Placebo or divalproex sodium in the treatment of manic or mixed episodes associated with bipolar 1 disorder
16. NCT00246246 - A Randomized, Open-label Trial of RISPERDAL® CONSTA™ Versus Oral Antipsychotic Care in Subjects With Bipolar Disorder

17. NCT00309699 - A Randomized, Double-Blind, Active- and Placebo-Controlled, Parallel-Group, Multicenter Study to Evaluate the Efficacy and Safety of Flexibly-Dosed, Extended-Release Paliperidone Compared With Flexibly-Dosed Quetiapine and Placebo in the Treatment of Acute Manic and Mixed Episodes Associated With Bipolar I Disorder
18. NCT00299715 - A Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Dose-Response, Multicenter Study to Evaluate the Efficacy and Safety of Three Fixed Doses of Extended-Release Paliperidone in the Treatment of Subjects With Acute Manic and Mixed Episodes Associated With Bipolar I Disorder
19. NCT00309686 - A Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Multicenter Study to Evaluate the Efficacy and Safety of Flexibly-Dosed Extended-Release Paliperidone as Adjunctive Therapy to Mood Stabilizers in the Treatment of Acute Manic and Mixed Episodes Associated With Bipolar I Disorder
20. NCT00490971 - A Randomized, Double-Blind, Active- and Placebo-controlled, Parallel-group, Multicenter Study to Evaluate the Efficacy and Safety of Extended-Release Paliperidone as Maintenance Treatment After an Acute Manic or Mixed Episode Associated With Bipolar I Disorder

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

## Research Proposal

## Project Title

Reevaluating Treatment Response Prediction in Mental Health: A Study on Generalizability

**Narrative Summary:**

A recent study using YODA Project data suggested that clinical prediction models may not generalize well beyond their development settings. Our proposal aims to investigate this further by reanalyzing those studies using different machine learning algorithms and techniques. We will also expand our focus beyond just one mental health disorder, schizophrenia (SZ), to include bipolar disorder (BD). By doing so, we aim to assess if validation issues are condition-specific. This research enhances predictive model development in healthcare, ultimately benefiting public health by ensuring more accurate diagnoses and treatments.

**Scientific Abstract:**

Background: Clinical prediction models are crucial tools in healthcare, yet their generalizability beyond specific settings is debated. Recent findings in SZ using YODA Project data suggest limitations in external validation, prompting further investigation.
Objective: Our study aims to reevaluate the generalizability of clinical prediction models by reanalyzing data from the same clinical trials using different machine learning algorithms and techniques. Beyond SZ, we seek to broaden the scope by including BD, to explore potential condition-specific validation challenges.
Study design: We will employ a retrospective analysis design, involving the reexamination of existing clinical prediction models developed for mental health disorders.
Participants: Individuals from trials in the YODA Project database, encompassing both SZ and BD populations, will be included.
Outcome measures: The primary outcome measure is the performance of clinical prediction models in SZ and in BD when validated externally, assessed through metrics such as accuracy, sensitivity, and specificity. Secondary outcomes include identifying factors influencing model generalizability and predictive power.
Statistical analysis: Various machine learning algorithms, including neural networks, decision trees,

and support vector machines, will be used to reanalyze existing data. Furthermore, subgroup analyses will be conducted to explore potential differences in model performance between SZ and BD cohorts. Recalibration techniques will also be applied to enhance model accuracy.

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

Clinical prediction models are pivotal in healthcare for informing treatment decisions based on statistical analyses of patient data. However, concerns persist regarding their applicability outside the specific settings where they were developed. Recent findings from the YODA Project suggest limitations in the external validation of these models, prompting a critical need for further investigation [1].
Our research project aims to address these concerns by reevaluating the generalizability of clinical prediction models, specifically focusing on mental health disorders. By reanalyzing previous studies utilizing diverse machine learning algorithms and techniques, we intend to assess the performance of these models across different populations and settings. Additionally, we aim to broaden the scope of our analysis by including bipolar disorder alongside schizophrenia, expanding beyond the traditionally studied condition. Beyond extending the work to another highly prevalent psychiatric disorder, we aim at investigating the efficacy of other types of machine learning models that have yet to be investigated. As seen in previous studies, the capabilities of machine learning models to generalize to new clinical trials poses a large challenge to the large-scale application of these approaches in clinical practice [1]. However, current studies have largely focused on linear models, such as the elastic net. We aim at investigating whether better models that are theoretically capable of modeling and generalizing much more complex phenomena, such as XGBoost [2], and other techniques such as recalibration [3], could address this issue and bring new ideas to the field of clinical translation of artificial intelligence.
This study holds significant importance for both science and public health. Firstly, it addresses a fundamental issue in predictive analytics within healthcare, contributing to the advancement of methodology in this field. By identifying potential limitations in model generalizability and exploring condition-specific challenges, our research will provide valuable insights into the factors influencing the effectiveness of clinical prediction models.
Furthermore, the outcomes of this study have direct implications for public health practice. Accurate clinical prediction models are essential for improving diagnostic accuracy, treatment efficacy, and patient outcomes. By enhancing our understanding of the strengths and limitations of these models, we can improve their utility in diverse healthcare settings, ultimately leading to more effective patient care and better public health outcomes.

**Specific Aims of the Project:**

Our primary aim is to reevaluate the performance of clinical prediction models within clinical trials, addressing concerns raised by previous research suggesting limitations in their generalizability. By applying alternative algorithms and techniques, such as recalibration, we hypothesize that these models will exhibit improved accuracy and performance even when externally validated. Key validation measures, including accuracy, sensitivity, and specificity, will be assessed to determine the effectiveness of these models across diverse populations and settings.
The secondary aim of our study is to expand the assessment of clinical prediction models to include bipolar disorders, in addition to schizophrenia. We aim to explore whether the difficulties encountered in validating prediction models for schizophrenia extend to BD. Our hypothesis posits that the models developed for BD will demonstrate superior performance compared to those for SZ. Clinical trials in BD were chosen when intervention had either risperidone or paliperidone, as they are most recommended for BD [4]. Through this comparative analysis, we aim to gain insights into the condition-specific challenges and potential differences in model performance across different mental health conditions.

**Study Design:**

Methodological research

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

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:**

We are seeking individual-level data from trials investigating the effects of antipsychotic treatments on SZ or BD. Within the YODA Project, we selected 5 trials on SZ to reevaluate data from a previous study [1], while we identified 15 trials to assess paliperidone and risperidone treatment [4] in BD. Our approach to inclusion and exclusion criteria are those established by each clinical trial. Our analysis will encompass participants of all ages diagnosed with SZ or BD based on the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV) criteria. To conduct these analyses effectively, we require access to individual-level data, including participant outcomes, treatment specifics, and personal characteristics necessary for adjusting for population differences (such as demographics, family history, and baseline mental health severity). We aim to include a wide range of participant demographics and characteristics precisely to maximize heterogeneity. The only specific exclusions will be individuals lacking complete information on primary outcome.

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

Internally and externally validated performance measures from treatment response predictive models will be evaluated. As performance measures, we will assess area under the ROC curve, accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and F1-score for the main classification models. Additional measures in supplementary regression-based models may be reported such as root mean squared error (RMSE), mean absolute error (MAE) and R2.
For SZ trials the primary outcome is change from baseline at 4 weeks measured by PANSS (Positive and Negative Syndrome Scale), while for BD trials we will consider change from baseline (number of weeks will be defined according to clinical trial data) measured by YMRS (Young Mania Rating Scale). The treatment response criteria may vary slightly among trials, but the focus is the difference in these endpoints from baseline to end of study. The treatment response will be input as a target for each model in order to estimate externally validated performance measures, this being considered the primary outcome. The secondary outcomes are the features that may influence model generalizability and predictive power.

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

Considering that we aim to develop predictive models for treatment response, the inclusion of diverse features (independent variables) is crucial to capture the underlying patterns within the data. Categorical data will be pre-processed as dummy variables. The following list provides potential predictors that could be included in our analysis: sex, race, ethnicity, age, BMI (body mass index), AIMS (Abnormal Involuntary Movement Scale) score, BARS (Barnes Akathisia Scale) score, CGI (Clinical Global Impression) severity, laboratory results, concomitant medication, diabetes and additional medical history, psychiatric history, time since diagnosis, electrocardiogram, family history, substance use, GAF (Global Assessment of Functioning Scale) score, smoking habit, hospitalization, MADRS (Montgomery-Asberg Depression Rating Scale) score, PANSS (Positive and Negative Syndrome Scale) score, physical exam, SARS (Simpson-Angus Scale) score, SF-36 (Short Form Health Survey) score, Sleep VAS Scale score, vital signs, YMRS score.

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

All digitally archived variables

**Statistical Analysis Plan:**

We will extend the analyses conducted by previous studies using state-of-the-art machine learning models and additional datasets [1] in bipolar disorders, due to positive results in machine learning for BD our group has found [5–9]. We aim at following the standard machine learning pipeline with a focus on creating generalizable models. First, we will only use two machine learning algorithms, elastic net and XGBoost. The elastic net will be useful for comparing results with previous studies, while the XGBoost will give us a better idea of the generalization of nonlinear models.
We will start the study by conducting a nested leave one out cross validation to optimize model hyperparameters (e.g. lambda for elastic net or learning rate for XGBoost) at the clinical trial level while maintaining an unbiased assessment. That means that each clinical trial will have two independent nested cross-validation, one for elastic net and one for XGBoost. Then, we will use all of the data from a single clinical trial to optimize the hyperparameters and train a model for each center. This model will then be used to predict the outcome on each of the other centers. By repeating this procedure for each trial, we will have results for each pair of clinical trials in both directions (once as training and once as testing). Finally, we will conduct a leave-one-trial-out approach, where we use training data from all but one clinical trial, testing the model on the trial that was left out. Thus, we will have a result for this experiment for each of the trials for when it was left as a test set.
We believe that a more thorough protocol including hyperparameter optimization and state-of-the-art nonlinear models will be enough to lead to generalized models. Although more generalized than a less fitted linear model, we still would expect a drop in accuracy when changing the data from one trial to the other. This problem is usually described as covariate shift or domain shift, meaning that there are distribution changes in the input data or changes in the relationship between input data and expected output that could lead to reduced performance [10]. This phenomenon has been investigated extensively in the field of semi-supervised learning and domain adaptation [10]. Evidence shows that reusing pretrained models that are trained in a different environment can drastically reduce the necessary data for training a model in a new environment. We will then extend our previous leave-one-trial-out approach by also introducing data points from the test trial into the training (thus removing them from the test set) and analyzing how that affects model performance. We will compare that with how many data points from that trial alone would be necessary to achieve a comparable performance.

**Software Used:**

Python

**Project Timeline:**

Project Start Date
Upon approval of the data request and receipt of access to the YODA Project Data. Estimated: March 1, 2024.

Analysis Completion Date
Three weeks after the project start date. Estimated: March 22, 2024.

Manuscript Drafted and Supplementary Materials Prepared
Following completion of data analysis, manuscript drafting and preparation of supplementary materials will commence. Estimated: March 23, 2024 - May 6, 2024.

Manuscript Submitted for Publication
One month after completion of manuscript drafting and preparation of supplementary materials. Estimated: June 6, 2024.

Results Reported Back to the YODA Project
Following publication of the manuscript and availability of results. Estimated: After publication,

anticipated August 2024.

**Dissemination Plan:**

The anticipated product of our project includes a manuscript aimed at publication in a high-impact scientific journal. This manuscript will detail our findings regarding the generalizability of clinical prediction models both in SZ and in BD. Specifically, we will produce one comprehensive article that presents our aims, hypotheses, methods, results, and implications. We anticipate that the manuscript will undergo rigorous peer review to ensure the validity and reliability of our findings. Ultimately, our goal is to publish novel insights to the field of clinical prediction modeling and advance scientific knowledge in mental health care.
Our target audience includes researchers, clinicians, and policymakers in psychiatry, psychology, data science, and public health. Stakeholders interested in predictive analytics and precision medicine in healthcare will also find our findings relevant.
Prestigious journals like Science, which published the paper on the illusory generalizability of prediction models, and Nature Medicine may be suitable for publication. These platforms offer broad readership and high impact, aligning with the interdisciplinary nature of our study.

**Bibliography:**

Chekroud AM, Hawrilenko M, Loho H, Bondar J, Gueorguieva R, Hasan A, et al. Illusory generalizability of clinical prediction models. Science. 2024;383: 164–167. doi:10.1126/science.adg8538

2. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: Association for Computing Machinery; 2016. pp. 785–794. doi: 10.1145/2939672.2939785
3. Riley RD, Ensor J, Snell KIE, Debray TPA, Altman DG, Moons KGM, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ. 2016;353: i3140. doi:10.1136/bmj.i3140
4. Yatham LN, Kennedy SH, Parikh SV, Schaffer A, Bond DJ, Frey BN, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) 2018 guidelines for the management of patients with bipolar disorder. Bipolar Disord. 2018;20: 97–170. doi:10.1111/bdi.12609
5. Passos IC, Mwangi B. Machine learning-guided intervention trials to predict treatment response at an individual patient level: an important second step following randomized clinical trials. Molecular psychiatry. 2020. pp. 701–702. doi:10.1038/s41380-018-0250-y
6. Passos IC, Mwangi B, Kapczinski F. Big data analytics and machine learning: 2015 and beyond. Lancet Psychiatry. 2016;3: 13–15. doi:10.1016/S2215-0366(15)00549-0
7. Passos IC, Ballester PL, Barros RC, Librenza-Garcia D, Mwangi B, Birmaher B, et al. Machine learning and big data analytics in bipolar disorder: A position paper from the International Society for Bipolar Disorders Big Data Task Force. Bipolar Disord. 2019;21: 582–594. doi: 10.1111/bdi.12828
8. Watts D, Pulice RF, Reilly J, Brunoni AR, Kapczinski F, Passos IC. Predicting treatment response using EEG in major depressive disorder: A machine-learning meta-analysis. Transl Psychiatry. 2022;12: 332. doi:10.1038/s41398-022-02064-z
9. Passos IC, Ballester P, Rabelo-da-Ponte FD, Kapczinski F. Precision Psychiatry: The Future Is Now. Can J Psychiatry. 2022;67: 21–25. doi:10.1177/0706743721998044
10. Farahani A, Voghoei S, Rasheed K, Arabnia, HR. A Brief Review of Domain Adaptation. arXiv (Cornell University). 2020 Oct 7