Principal Investigator

First Name: Anthony
Last Name: Barrows
Degree: MS
Primary Affiliation: University of Vermont
E-mail: ajbarrow@uvm.edu
State or Province: Vermont
Country: USA

General Information

Key Personnel (other than PI):
First Name: Nicholas
Last name: Allgaier
Degree: PhD
Primary Affiliation: University of Vermont
SCOPUS ID:
Requires Data Access? Yes

First Name: Hugh
Last name: Garavan
Degree: PhD
Primary Affiliation: University of Vermont
SCOPUS ID:
Requires Data Access? Yes

Are external grants or funds being used to support this research?: External grants or funds are being used to support this research.
Project Funding Source: NIDA T32DA045593
How did you learn about the YODA Project?: Scientific Publication

Conflict of Interest

https://yoda.yale.edu/wp-content/uploads/2024/04/YODA_COI_AJB.pdf
https://yoda.yale.edu/wp-content/uploads/2024/04/YODA_COI_NA.pdf
https://yoda.yale.edu/wp-content/uploads/2024/04/YODA-COI-Garavan.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. 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
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

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

Research Proposal

Project Title

Robust hyperparameter selection needed to build generalizable clinical prediction models

Narrative Summary:

A previous study by Chekroud et al. [1] found limitations of out-of-study predictive power of machine learning algorithms using five studies on schizophrenia treatment outcomes (NCT00518323, NCT00334126, NCT00085748, NCT00650793, and NCT00083668). Specifically, the authors used several modeling paradigms to predict dichotomized schizophrenia remission and compared each model's performance using balanced accuracy. They found that many models which showed strong in-sample fit performed poorly out of sample and concluded that models predicting schizophrenia treatment outcomes may have limited generalizability.

Although the authors employed robust methods to estimate out-of-sample model performance (i.e., k-fold cross validation), their methods may fail to adequately distinguish between cross validation for effect size estimation and cross validation for model selection. We submit that estimating hyperparameters using a nested cross-validation scheme is critical for constructing generalizable models in the first place.

Scientific Abstract:

Background:
Statistical models used to predict clinical outcomes are most useful when they perform equally well in cohorts other than the ones in which they were fit. However, a recent study found limitations in the generalizability of models meant to predict likely schizophrenia treatment outcomes [1].

Objective:
This study adds to clinical outcomes prediction literature by implementing a model-fitting procedure designed to prevent data leakage, fit generalizable models, and obtain accurate estimates of out-of-sample predictive performance.

Study Design:
Several multivariate predictive models will be used to predict early and late-responders to schizophrenia treatment using clinical and demographic attributes as independent variables.

Participants:
Participants from five, international clinical trials of paliperidone to treat schizophrenia will be included in this analysis (n=1962).

Primary and Secondary Outcome Measure(s):
The main outcome will be balanced accuracy from each model predicting schizophrenia treatment outcomes.

Statistical Analysis:
Four modeling scenarios will be implemented to illustrate the performance of clinical prediction models, each employing elastic net regression and a nested cross-validation scheme.

Brief Project Background and Statement of Project Significance:

Model generalizability, or the extent to which a model which predicts a clinical outcome in one cohort survives when the model is applied to another, can have critical implications for patient care [2]. Although a model's prediction performance on a new, independent population sample cannot be known, it can be reliably estimated.

Stone [3] distinguished between cross-validatory techniques used for estimating model effect size and those used for selecting optimal model hyperparameters in 1974. Several different implementations of this procedure have appeared in the literature e.g., [4], [5], [6], although each algorithm involves an "outer" effect size validation loop combined with an "inner" loop for model selection. ¬¬¬¬¬

Although these techniques have been correctly described as more computationally expensive than their so-called "flat" cross validation counterparts (i.e., where model parameters are tuned concurrently with effect size estimation) [7], [8], the increasing availability of vast computing power allows researchers to easily navigate this expense in favor of avoiding overfit models [5]. Flat cross validation procedures require selecting median hyperparameter values, meaning these selections are not optimal for any of the models fit to dataset partitions used for evaluation. Furthermore, because these median values are estimated using the whole dataset, subsequent data partitions are no longer statistically independent, presenting a problem of data leakage [7].

Notably, this model-fitting procedure is often not implemented in popular predictive modeling software packages, including caret [9], used in [1]. Consequently, one reason for the limited use of nested validation procedures is likely that researchers would need to implement these techniques themselves, or to rely on implementations which may not be compatible with their preferred statistical analysis software. Nevertheless, nested validation algorithms are relatively straightforward to implement in object-oriented programming languages like Python, or R (see [10] for a purpose-built library).

The present work builds on Chekroud et al. [1] work predicting schizophrenia treatment outcomes by repeating their analyses using a nested cross validation framework. In order to assess the differences in out-of-sample effect size estimates directly attributable to model fitting and evaluation, the current analysis adheres to that of the original authors' as closely as possible. Specifically, nested cross validation is performed using caret [9] in R, and effect size estimates are compared with those from the original authors' analyses. All code will be publicly available.

Specific Aims of the Project:

This project aims to demonstrate the importance of using nested cross-validation procedures for training and evaluating multivariate predictive models to predict clinical outcomes. Specifically, we aim to show that rigorously trained models display similar in-sample and out-of-sample performance, even when applied to an unseen data set. We will demonstrate this performance using models predicting treatment response to paliperidone in patients with schizophrenia.

Study Design:

Methodological research

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

Develop or refine statistical methods

Research Methods

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

We will include participants enrolled in the randomized controlled trials from NCT00518323, NCT00334126, NCT00085748, NCT00078039, and NCT00083668. All participants will have a diagnosis of schizophrenia at the start of the trial. All participants who completed their trial (i.e., have outcome data available at week 4) will be included. There are no explicit exclusion criteria.

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

The outcome of interest will be balanced accuracy of multivariate predictive models predicting treatment response using the PANSS [11] using the Remission in Schizophrenia Working Group (RSWG) criteria [12]. Secondary outcomes will be alternative definitions of schizophrenia treatment outcomes, including 25% and 50% reductions in PANAS score, and total percent change in PANAS total score, corrected for baseline PANAS.

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

Standardized dose of paliperidone will be used as the primary predictor, using dose equivalents as necessary [13].

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

Elastic net regression will be used to select a subset of baseline characteristics which are available for more than 80% of participants. In particular, we will include sociodemographic variables (i.e., sex, race, age, DSM IV type of schizophrenia diagnosis, age of diagnosis, and number of hospitalizations), blood and urinalysis data, and other schizophrenia symptoms scales (i.e., Abnormal Involuntary Movement Scale, Simpson Angus Scale) and treatment randomization.

Statistical Analysis Plan:

We will use regularized regression (i.e., elastic net) to identify a set of uncorrelated or minimally correlated independent variables to predict schizophrenia treatment outcomes (RSWG criteria). Crucially, we will fit these predictive models using a nested cross validation scheme consisting of “outer” and “inner” validation loops. In the outer loop, the dataset will be divided into 10 folds, with one statistical model fit per fold. To determine optimal hyperparameters for these models (i.e., alpha and lambda), each outer fold will be further divided into 10 inner folds, from which median hyperparameters will be selected for the model fit to the outer fold. Out-of-sample effect sizes will be estimated using the outer-loop models and reported using balanced accuracy.

We will implement the methods described above in four modeling schemes designed to assess whether training generalizable models to predict clinical outcomes is feasible using these data sets. We will (1) assess predictive performance within trials without cross-validation, (2) repeat this procedure using cross-validation, (3) use paired-across-trial prediction (i.e., train the model on one

trial, evaluating its performance using data from the others), and (4) use leave-one-out cross validation, training a model on four trials and evaluating its performance using the fifth.

Software Used:

R

Project Timeline:

We plan to initiate this project upon receipt of the data. Presuming the requested data are made available to us June 1, 2024, we expect results by August 1, 2024, and will draft a report for September 1, 2024. Results will be reported back to the YODA Project in accordance with the Data Use Agreement.

Dissemination Plan:

We will submit our report for publication in the journal Science. Specifically, we aim to enhance the literature concerning generalizable predictive models for clinical outcomes, which we believe is of interest to a wide audience.

Bibliography:

1] A. M. Chekroud *et al.*, “Illusory generalizability of clinical prediction models,” *Science*, vol. 383, no. 6679, pp. 164–167, Jan. 2024, doi: 10.1126/science.adg8538.

[2] H. Singh, V. Mhasawade, and R. Chunara, “Generalizability challenges of mortality risk prediction models: A retrospective analysis on a multi-center database,” *PLOS Digit Health*, vol. 1, no. 4, p. e0000023, Apr. 2022, doi: 10.1371/journal.pdig.0000023.

[3] M. Stone, “Cross-Validatory Choice and Assessment of Statistical Predictions,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 36, no. 2, pp. 111–133, Jan. 1974, doi: 10.1111/j.2517-6161.1974.tb00994.x.

[4] S. Varma and R. Simon, “Bias in error estimation when using cross-validation for model selection,” *BMC Bioinformatics*, vol. 7, no. 1, p. 91, Dec. 2006, doi: 10.1186/1471-2105-7-91.

[5] D. Krstajic, L. J. Buturovic, D. E. Leahy, and S. Thomas, “Cross-validation pitfalls when selecting and assessing regression and classification models,” *J Cheminform*, vol. 6, no. 1, p. 10, Dec. 2014, doi: 10.1186/1758-2946-6-10.

[6] N. Iizuka *et al.*, “Oligonucleotide microarray for prediction of early intrahepatic recurrence of hepatocellular carcinoma after curative resection,” *The Lancet*, vol. 361, no. 9361, pp. 923–929, Mar. 2003, doi: 10.1016/S0140-6736(03)12775-4.

[7] G. C. Cawley and N. L. C. Talbot, “On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation”.

[8] J. Wainer and G. Cawley, “Nested cross-validation when selecting classifiers is overzealous for most practical applications.” arXiv, Sep. 25, 2018. Accessed: Apr. 08, 2024. [Online]. Available: http://arxiv.org/abs/1809.09446

[9] M. Kuhn, “Building Predictive Models in *R* Using the **caret** Package,” *J. Stat. Soft.*, vol. 28, no. 5, 2008, doi: 10.18637/jss.v028.i05.

[10] M. J. Lewis, A. Spiliopoulou, K. Goldmann, C. Pitzalis, P. McKeigue, and M. R. Barnes, “nestedcv: an R package for fast implementation of nested cross-validation with embedded feature selection designed for transcriptomics and high-dimensional data,” *Bioinformatics Advances*, vol. 3, no. 1, p. vbad048, Jan. 2023, doi: 10.1093/bioadv/vbad048.

[11] S. R. Kay, A. Fiszbein, and L. A. Opler, “The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia,” *Schizophrenia Bulletin*, vol. 13, no. 2, pp. 261–276, Jan. 1987, doi: 10.1093/schbul/13.2.261.

[12] N. C. Andreasen, W. T. Carpenter, J. M. Kane, R. A. Lasser, S. R. Marder, and D. R. Weinberger, “Remission in Schizophrenia: Proposed Criteria and Rationale for Consensus,” *AJP*, vol. 162, no. 3, pp. 441–449, Mar. 2005, doi: 10.1176/appi.ajp.162.3.441.

[13] S. Leucht, M. Samara, S. Heres, and J. M. Davis, “Dose Equivalents for Antipsychotic Drugs: The DDD Method: Table 1.,” *SCHBUL*, vol. 42, no. suppl 1, pp. S90–S94, Jul. 2016, doi: 10.1093/schbul/sbv167.