Paper ML02

Streamlining Machine Learning Modeling in Neuroscience Clinical Trials through Cross-Functional Collaboration

Jing Dai, Jazz Pharmaceuticals and Wayne State University, Detroit, MI, USA
Brian Shonesy, Jazz Pharmaceuticals, Philadelphia, PA, USA
Jiayu Hu, Northwestern University, Chicago, IL, USA
Ming Qiang Huang, Jazz Pharmaceuticals, Philadelphia, PA, USA

ABSTRACT

Neuroscience clinical trials face common challenges including high placebo responses and high attrition rates. Machine learning (ML) effectively identifies relevant baseline characteristics to inform future trial designs. Unlike traditional statistical methods, ML techniques offer advantages such as handling high dimensionality and performing automated model tuning, leading to more robust and accurate predictive models.

This paper focuses on streamlining ML modeling using neuroscience clinical trial data in a cross-functional setting. Key steps include (1) creating a ML-ready data file from ADaM datasets by clinical biostatisticians or programmers, (2) consulting clinicians for baseline feature pre-selection, (3) training, assessing, and fine-tuning ML models by data scientists, and (4) summarizing, visualizing, and interpreting ML results in collaboration with clinicians. This paper will also cover methodologies to enhance model fitness and to prevent overfitting, when dealing with small cohorts, significant missing data, and a large number of features—typical obstacles in modeling clinical trial data.

INTRODUCTION

Neuroscience clinical trials face multifaceted challenges that can significantly impact the development of novel therapeutics. Two of the most pressing issues are high placebo response rates and elevated participant attrition. These issues can obscure treatment effects, reduce statistical power, and ultimately lead to false negative results, potentially hindering the development of promising interventions.

Placebo response rates in neuroscience trials, particularly in areas such as depression, pain, and Alzheimer's disease, can be substantial and pose significant challenges to drug development. Studies have consistently shown that placebo response rates may be as high as 30%, and in some cases, can be even higher [1]. For instance, in major depressive disorder trials, placebo response rates have been reported as high as 50% [2]. This high placebo effect can mask treatment efficacy, leading to failed trials for potentially effective compounds.

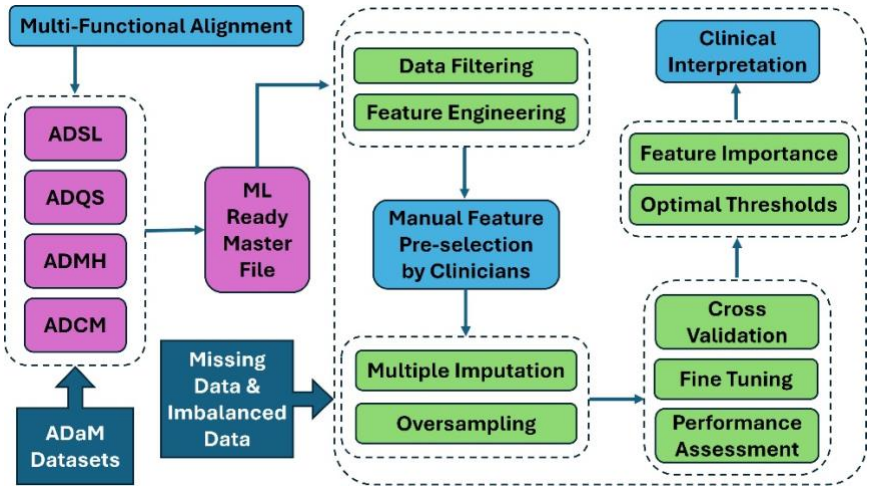
Attrition rates in neuroscience clinical trials can exceed 30% [3]. A review of 71 randomized controlled trials published in top medical journals found attrition rates of more than 20% in 18% of the trials [4]. High dropout rates are particularly problematic in long-term studies, such as those for chronic neurological or psychiatric conditions. For example, in Alzheimer's disease trials, dropout rates have been reported to range from 20% to 40% over 12 to18 month periods [5]. These high attrition rates can lead to missing data, potentially biasing results and compromising the validity and generalizability of study findings.

In response to these challenges, the field has turned to advanced analytical methods, particularly machine learning (ML) approaches. These sophisticated algorithms offer the potential to identify complex patterns in baseline characteristics that contribute to high placebo response and attrition rates, which may inform future study design to minimize the impact of these challenges. Despite the typically small cohort sizes in clinical trials, ML models can capture non-linear relationships and provide insights beyond those of traditional statistical methodology. This paper presents a standardized, cross-functional approach to the implementation of ML modeling in neuroscience clinical trials, specifically tailored to handle the constraints of limited cohort sizes. The proposed framework ensures data consistency and reliability across different stages of analysis, from initial data preparation to final interpretation, while employing techniques suited for small datasets. Moreover, it maximizes interdisciplinary collaboration, integrating expertise from statistical programmers, biostatisticians, clinicians, and data scientists. By adopting this systematic methodology, researchers can enhance their ability to predict and mitigate factors contributing to high placebo response and attrition rates, even with limited data, ultimately informing more robust and efficient trial designs in the challenging landscape of neuroscience drug development. Additionally, future applications of this work include the development of new predictive assessments tools to support trial design and decision-making by forecasting treatment response or dropout likelihood. An interactive patient profiling application could be created to help clinicians visualize participant profiles, including placebo risk

and dropout probability, facilitating more informed clinical decisions. Precision treatment strategies can be advanced by identifying subgroups of patients most likely to benefit, thereby minimizing adverse effects and placebo responses. These tools and applications could also be integrated into ongoing clinical trials, enabling real-time risk assessment and monitoring to improve trial efficiency.

WORKFLOW

Figure 1. Workflow of machine learning modeling in neuroscience clinical trials.



The flowchart (Figure 1) outlines a structured approach for implementing ML modeling in neuroscience clinical trials. Below is a step-by-step breakdown of the process:

- **Pre-modeling alignment and documentation:** The pre-modeling alignment phase typically begins with a kick-off meeting involving clinicians, clinical biostatisticians, statistical programmers, and data scientists to ensure that all team members are aligned on the modeling objectives. This meeting establishes a shared understanding of the source data, modeling scope, foreseeable risks and challenges, project timelines, and each team member's roles and responsibilities. Documenting a project synopsis and a detailed modeling plan is highly recommended to facilitate and solidify this alignment. Such preparation is critical for effective collaboration and smooth progress throughout the project.

- **Data integration:** After these initial discussions, clinical biostatisticians or programmers create an ML-ready master dataset that encompasses all relevant baseline characteristics, including subject-level data, baseline questionnaire assessment scores, medical history, concomitant medication, and other necessary datasets. This step emphasizes data comprehensiveness, consistency, and reliability.

- **Feature engineering and manual feature pre-selection:** Once the ML-ready master dataset is created, the next step involves careful feature engineering and manual feature pre-selection. This process begins with data filtering, where decisions are made to include only the relevant data for accurate analysis—such as selecting study arms, study periods, and analysis sets—based on the modeling objectives. Feature engineering then prepares the data for modeling by recoding categorical or string variables into numerical formats, creating new features, modifying existing ones to enhance model interpretability and performance, and removing features with scarce data. Following these preparatory steps, clinicians manually review the feature list and pre-select features, identifying variables with the most clinical relevance. This ensures that the modeling process is both grounded in clinical insights and optimized for meaningful outcomes.

- **Multiple imputation and oversampling:** With engineered and pre-selected features in place, techniques such as multiple imputation and oversampling are applied to address remaining issues with missing data and class imbalance, when necessary. Multiple imputation (MI) imputes missing values while introducing variability, ensuring the imputed data reflects more realistic patterns. Oversampling then adjusts the dataset to provide adequate representation for underrepresented classes. Together, these techniques enhance data quality and improve the robustness and generalizability of the model. However, these techniques are not without limitations. They may introduce biases or noise

into the data, which could affect the model's performance and generalizability in unseen scenarios. Strategies to mitigate these risks, such as sensitivity analyses and external validation, will be explored in the **Discussion** section.

- **Model training, validation, and optimization:** Once the refined dataset is ready, ML models are trained to identify baseline characteristics that predict placebo response or participant attrition. Before training, preprocessing steps such as standardization (e.g., z-score normalization) and scaling (e.g., Min-Max scaling) are applied to ensure that all features are on a comparable scale, which is particularly important for algorithms sensitive to feature magnitudes, such as support vector machines (SVMs) and logistic regression.

  The process begins with model training, during which the data is split into training and validation subsets, often using k-fold cross-validation (CV) with 5 to 10 folds to reduce variability in performance estimates. Grid search or randomized search is employed during CV to systematically explore and identify the optimal hyperparameters, ensuring the model performs well on the training set while avoiding overfitting. Common useful packages for this step include scikit-learn (Python) or caret/tidymodels (R).

  After identifying the best parameters, fine-tuning is conducted on a narrower hyperparameter space to further optimize the model's performance. During this phase, advanced techniques such as Bayesian optimization or automated machine learning (AutoML) can be used to refine the model more efficiently.

  For performance assessment, the model's effectiveness is evaluated on the test set to measure its generalizability. Depending on the outcome type. For binary outcomes, metrics such as accuracy, precision, recall, F1-score, ROC-AUC, and PR-AUC are used to assess classification performance. For continuous outcomes, metrics such as mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and R-squared are reported. Visual tools, such as ROC curves for binary outcomes or residual plots for continuous outcomes, can be employed to communicate model performance effectively. This iterative process of training, validation, and optimization ensures the model is robust, generalizable, and capable of delivering clinically meaningful insights.

- **Feature importance and optimal thresholds:** After assessing the model's performance and confirming it meets the necessary standards; feature importance and optimal thresholds are determined. Feature importance analysis highlights the most influential variables within the model, providing insights into which baseline characteristics significantly impact placebo response or attrition. Optimal thresholds are then identified to balance sensitivity and specificity according to the study's goals, such as establishing cutoff values for inclusion and exclusion criteria in future trials. This step enhances the interpretability and practical applicability of the model's findings, making the results more actionable for clinical trial design.

- **Clinical Interpretation:** Clinicians review the modeling outputs to assess its relevance and applicability to clinical practice. This stage integrates the modeling results with clinical expertise, providing insights that can guide decision-making in the context of neuroscience clinical trials. Visualizations play a key role in this step, helping to present complex model findings in a more interpretable format, which aids clinicians in understanding and applying the results effectively.

Importantly, the steps in this workflow are highly iterative, particularly in feature engineering, model optimization, and clinical interpretation. Continuous adjustments and refinements are necessary as insights from each stage feed back into earlier steps, helping to fine-tune model parameters. This iterative approach is essential for achieving a model that not only meets performance standards but also holds clinical relevance, ensuring that the findings are both statistically robust and practically applicable for improving future trial design and outcomes.

DISCUSSIONS

The overall objective of these ML modeling efforts in neuroscience clinical trials is to identify impactful baseline characteristic variables and the associated optimal threshold values that can inform future trial designs and support clinical development programs. By pinpointing these predictive factors, we aim to better understand and manage placebo response and participant attrition, ultimately enhancing the efficiency of future neuroscience clinical trials.

The key reason for choosing ML techniques over traditional statistical models, such as linear and logistic regression with forward or backward selection, lies in ML's capacity to uncover non-linear relationships and handle high-dimensional data in neuroscience clinical trials. Among ML techniques, elastic net regression is particularly advantageous and widely used in clinical trials because it combines the benefits of both ridge and lasso regression, directly reporting regression coefficients while managing multicollinearity and feature selection effectively. Elastic net is well-suited for identifying predictive relationships within a large set of baseline characteristics, which often include numerous and potentially correlated variables. This technique, along with other ML methods like random forests and gradient boosting, excels at capturing non-linear relationships and interactions

that traditional statistical models may miss. By leveraging the flexibility of ML, researchers can achieve more accurate and clinically meaningful insights, enhancing trial design and helping to identify baseline predictors for improved participant management.

Under this context, several unique challenges warrant discussion. These include the complexity of high dimensionality relative to the typically small cohort sizes and the need to address missing and/or imbalanced data effectively. While methodologies exist to enhance model fitness and prevent overfitting in such scenarios, they often require adaptation, refinement, or integration with manual processes and human expertise.

For example, while dimensionality reduction methods such as Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE) can mitigate overfitting, they raise interpretability issues that are critical in medical research [7]. In clinical contexts, each feature represents a specific biomarker, symptom, or patient characteristic with direct medical relevance, and transforming these into abstract or non-linear components—as in PCA or t-SNE—can obscure their meaning. This loss of interpretability complicates translating findings into actionable clinical insights and may create challenges with regulatory compliance. To address this, we implemented a combination of manual feature pre-selection and model-based techniques such as Recursive Feature Elimination (RFE) to prioritize variables that are both clinically meaningful and statistically impactful. Engaging a group of clinicians in feature pre-selection further reduces individual bias and ensures the inclusion of relevant variables, anchoring the model in clinical insights while retaining interpretability—an essential requirement for effective and transparent applications in clinical trials.

Addressing missing and/or imbalanced data also requires thoughtful adaptation of methods to ensure robust and interpretable results. MI is a powerful approach for handling missing data, especially in small cohorts. By generating multiple datasets where missing values are imputed based on relationships among observed variables, it allows ML models to retain statistical power and minimizes the loss of valuable information. Voting methods and ensemble techniques are also available for summarizing feature selection and importance following imputation. However, MI can introduce bias if the imputation model is mis-specified or if incorrect assumptions about the missing data mechanism are made. Moreover, MI can be computationally expensive, particularly with high-dimensional data and iterative model training. Similarly, oversampling methods such as Synthetic Minority Oversampling Technique (SMOTE) and Adaptive Synthetic Sampling (ADASYN) are effective for addressing class imbalances by generating synthetic samples for underrepresented classes. Despite their utility, these methods can introduce bias and noise, increase the risk of overfitting to synthetic samples, and complicate decision boundaries, leading to more misclassifications. To maintain transparency and reliability, biases and noise introduced by these manipulations must be clearly communicated with the clinical team. Methods to quantify and mitigate these effects, such as sensitivity analysis, validation on external datasets, and Shapley value analysis for bias assessment, should be actively integrated into the workflow to ensure reliable results.

Last but not least, efficient and effective communication is key to the success of ML projects in neuroscience clinical trials through cross-functional collaboration. Ensuring a multifunctional team stays aligned and engaged throughout the project requires deliberate communication and educational strategies. Regular meetings are vital to provide updates, address challenges, and refine objectives, fostering a shared understanding of progress and goals. To enhance the team's expertise and engagement, educational materials such as a knowledge slide deck and a cheat sheet can be invaluable. These materials should include key ML concepts, important model parameters, and performance assessment metrics, allowing team members to quickly familiarize themselves with essential topics. Simplifying technical details and incorporating project-relevant examples make these resources accessible to all team members, regardless of their technical background. Encouraging active participation, open questions, and collaborative problem-solving during meetings further strengthens the team's cohesion.

REFERENCES

1. Walsh, B. T., Seidman, S. N., Sysko, R., & Gould, M. (2002). Placebo response in studies of major depression: variable, substantial, and growing. JAMA, 287(14), 1840-1847.
2. Furukawa, T. A., et al. (2016). Placebo response rates in antidepressant trials: a systematic review of published and unpublished double-blind randomised controlled studies. The Lancet Psychiatry, 3(11), 1059-1066.
3. Wahlbeck, K., Tuunainen, A., Ahokas, A., & Leucht, S. (2001). Dropout rates in randomised antipsychotic drug trials. Psychopharmacology, 155(3), 230-233.
4. Wood, A. M., White, I. R., & Thompson, S. G. (2004). Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. Clinical Trials, 1(4), 368-376.
5. Grill, J. D., & Karlawish, J. (2010). Addressing the challenges to successful recruitment and retention in Alzheimer's disease clinical trials. Alzheimer's Research & Therapy, 2(6), 34.
6. Smith, E. A. et al. (2022). Using artificial intelligence-based methods to address the placebo response in clinical trials. Innovations in Clinical Neuroscience, 19(1–3), 60–70.

7. Jolliffe IT, Cadima J. (2016). Principal component analysis: a review and recent developments. Phil. Trans. R. Soc. A 374:20150202.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Jing Dai
Director Biostatistics
Jazz Pharmaceuticals
Adjunct Faculty
Department of Mathematics, Wayne State University.
Email: jdai@jazzpharma.com