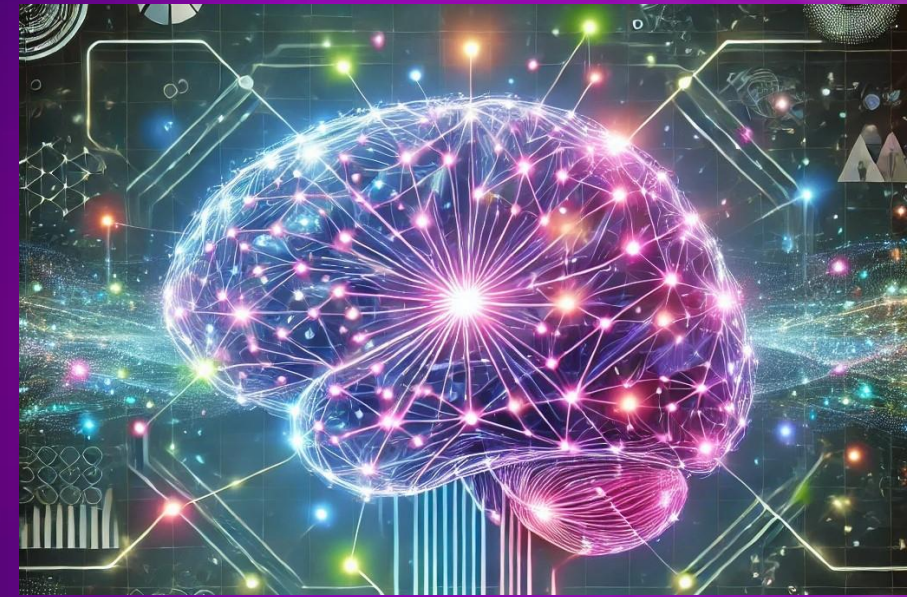

Streamlining Machine Learning Modeling in **Neuroscience Clinical Trials** through Cross-Functional Collaboration



Jing Dai
Director Biostatistics
Jazz Pharmaceuticals

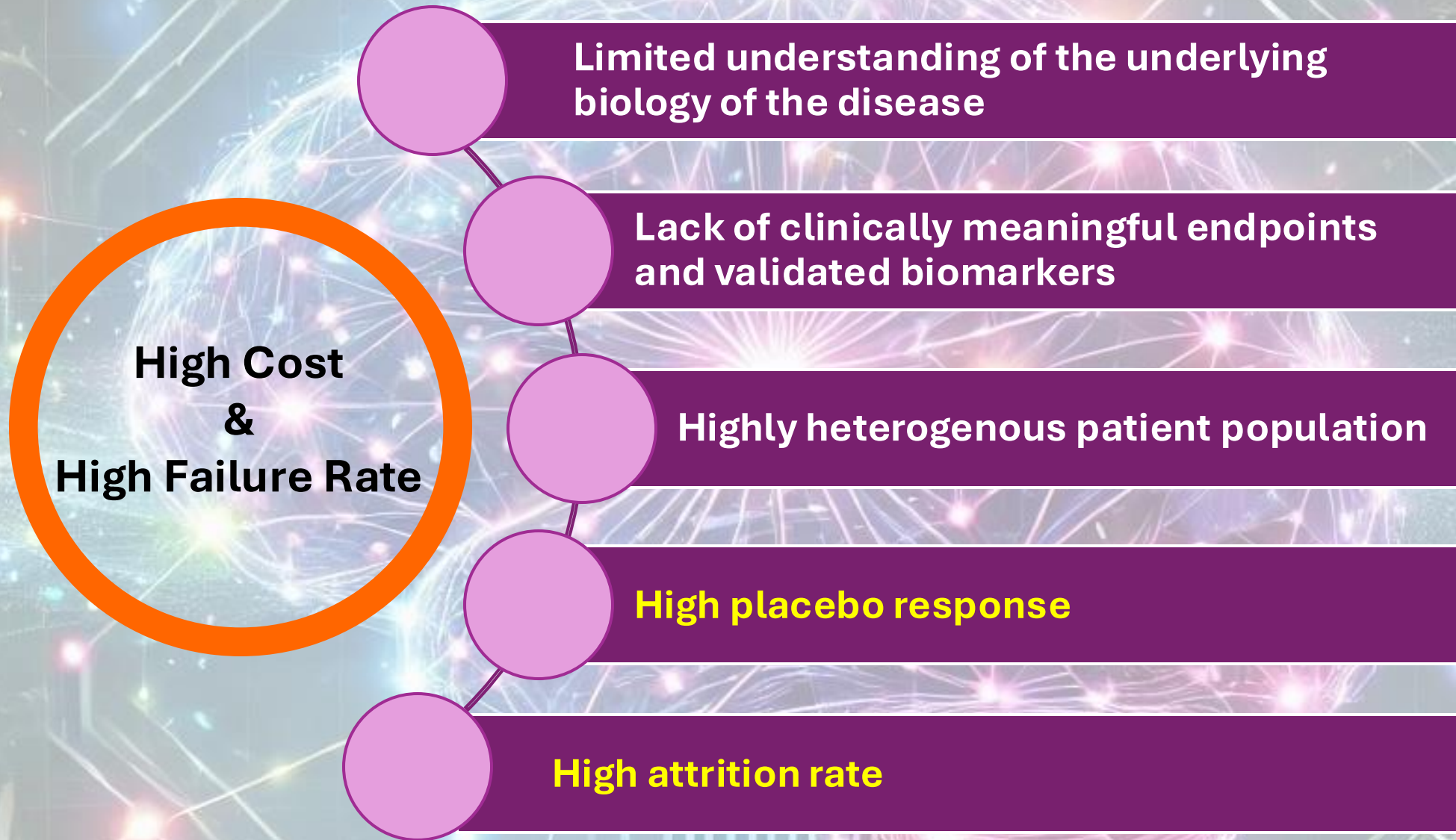
PHUSE US CONNECT, March 16-19, 2025



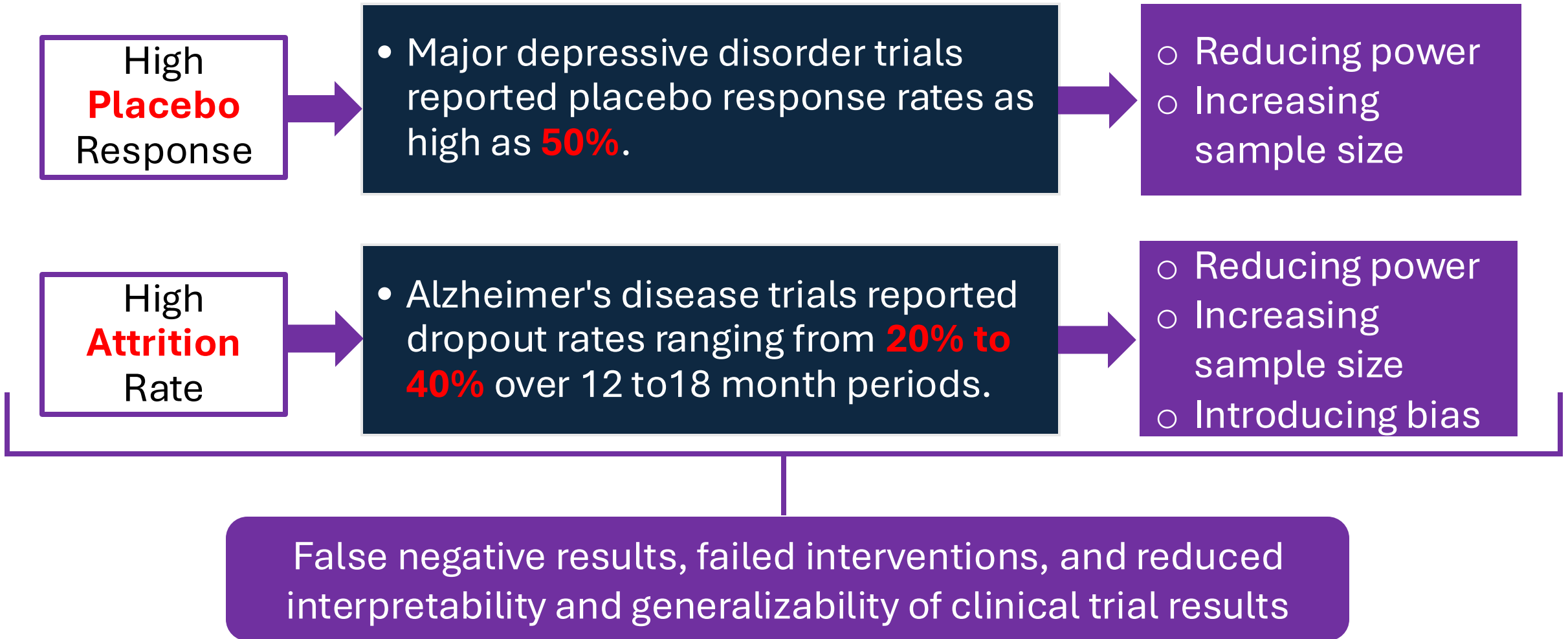
Disclaimer

The views and opinions expressed in this presentation are those of the speaker and do not necessarily reflect the positions of any entities they represent.

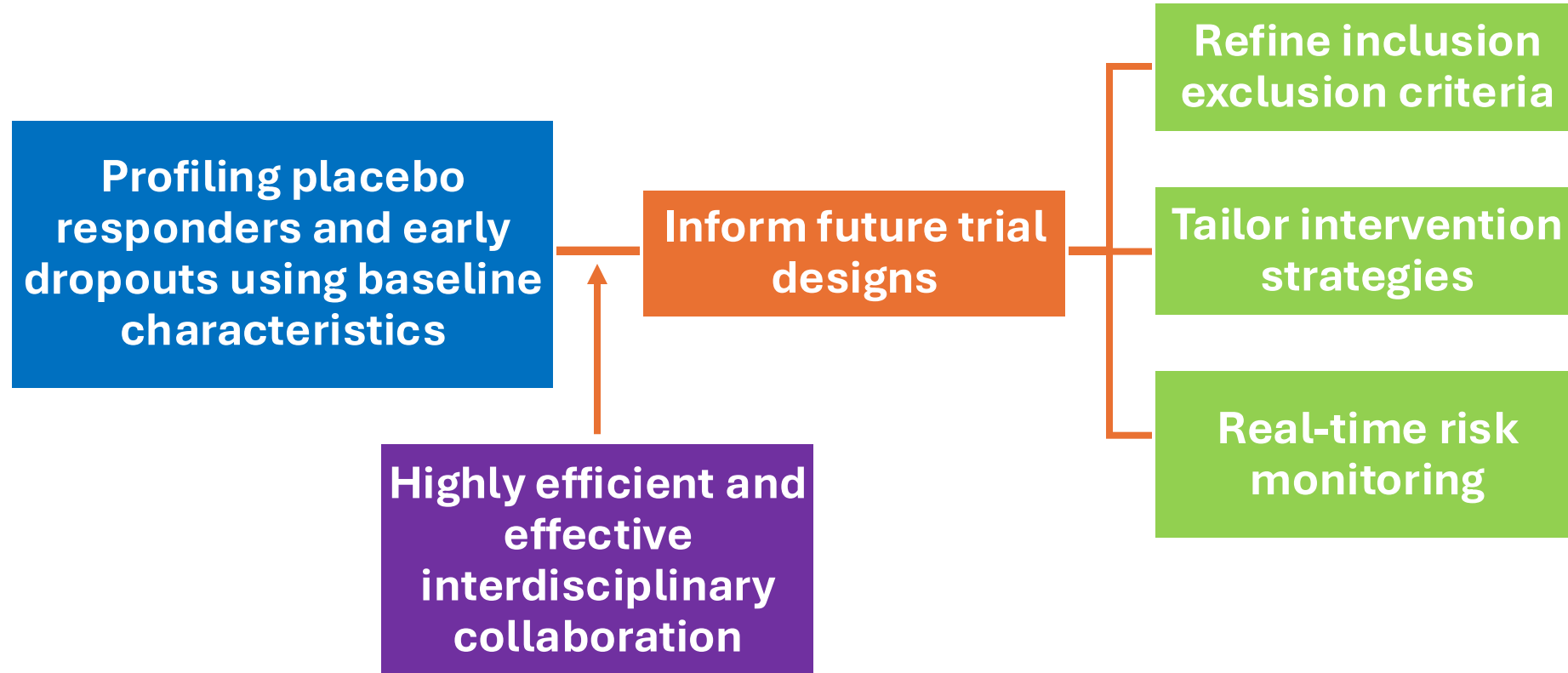
Challenges of Neuroscience Clinical Trials



Challenges of Neuroscience Clinical Trials



Modeling Objectives



VS

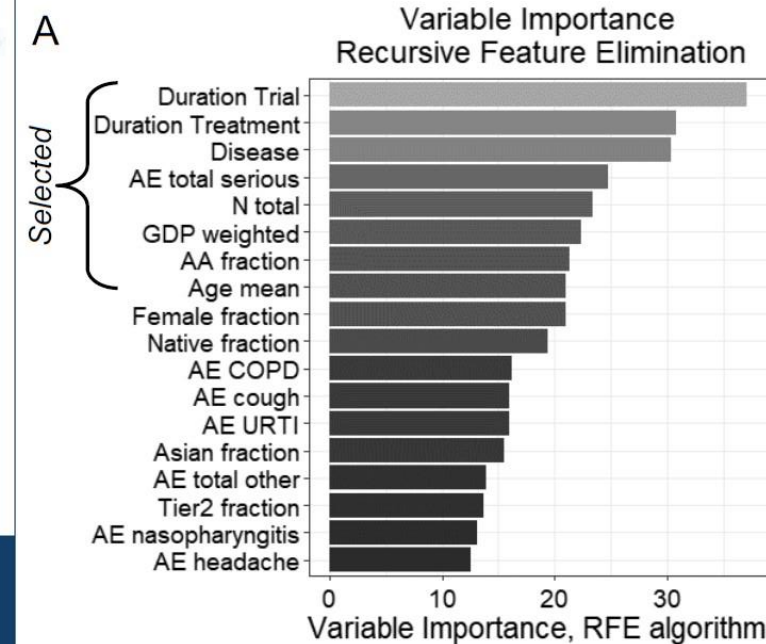
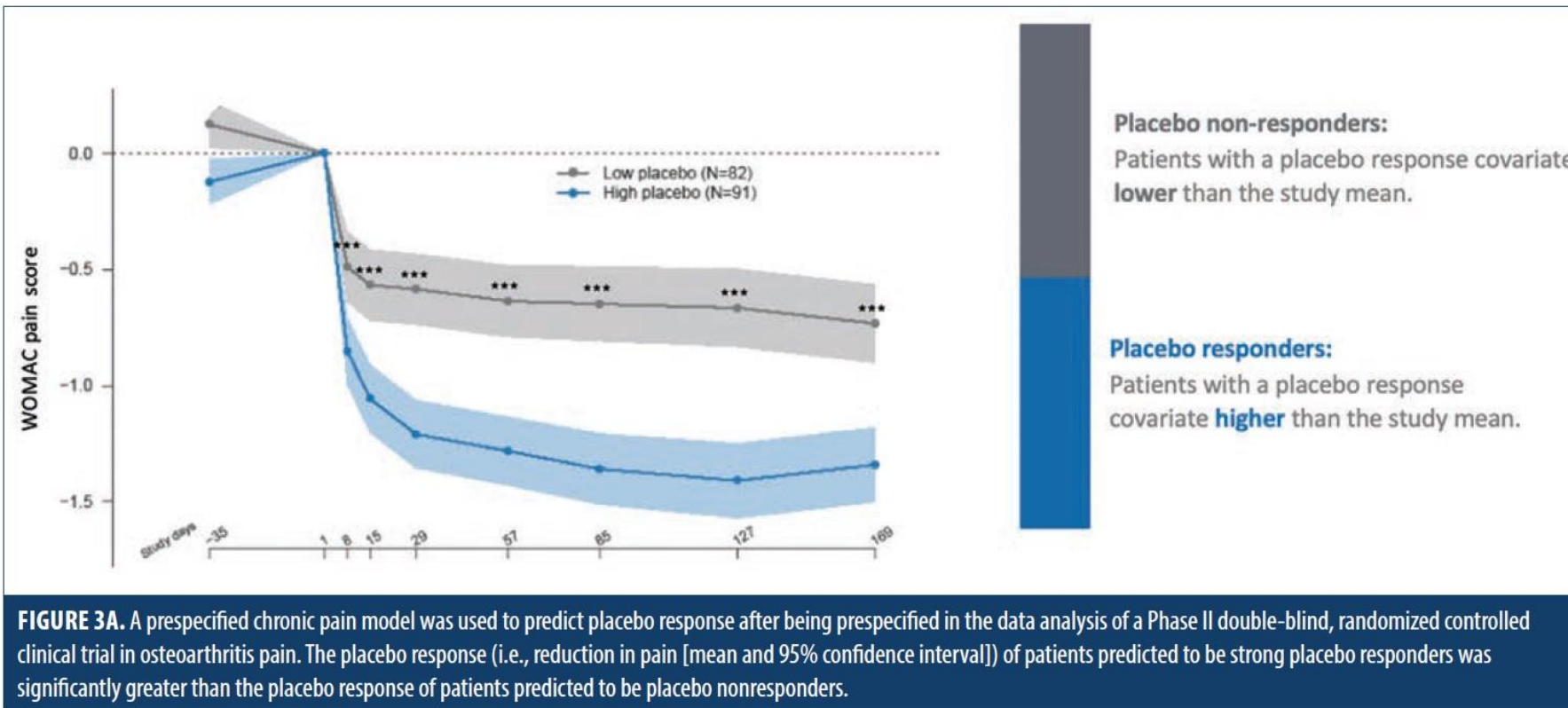
Machine Learning

- ❖ Handles **non-linear** relationships.
- ❖ Excels in **high-dimensional** data.
- ❖ Can manage **small cohort sizes** while still providing robust insights.
- ❖ **Automatically** performs feature selection and model optimization.
- ❖ Captures **complex patterns** that traditional methods may overlook.

Traditional Statistics Modeling

- ❖ Assumes **linear** relationships.
- ❖ Struggles with **high-dimensional data** due to multicollinearity.
- ❖ Less effective with **small sample sizes**.
- ❖ Requires **manual** feature selection and predefined model assumptions.
- ❖ May miss **complex non-linear interactions** between variables.

Machine Learning Modeling Examples



KOM

Data
Integration

Data
Engineering

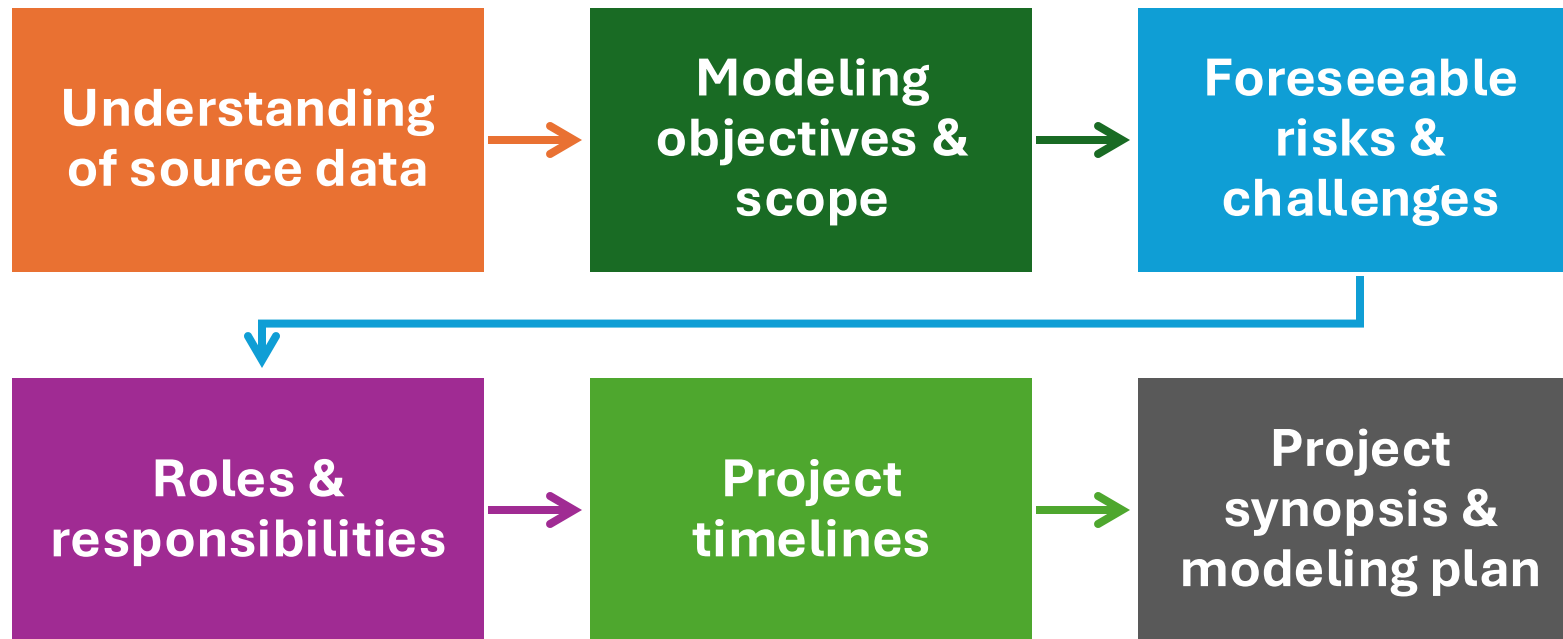
Modeling

Reporting

Interpretation

1

KOM: Pre-modeling multi-functional alignment and documentation



KOM

Data
Integration

Data
Engineering

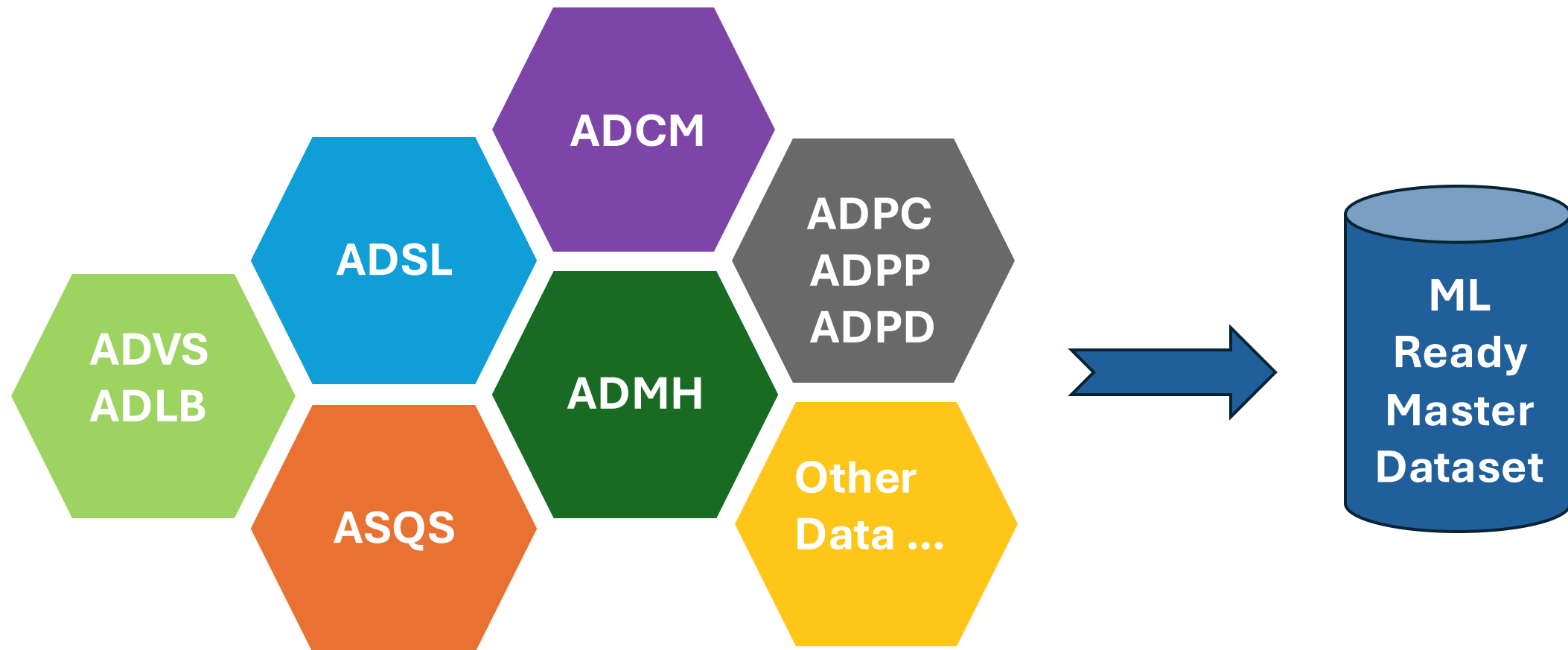
Modeling

Reporting

Interpretation

2

Data Integration: Assemble the machine learning ready master data file



3

Data Engineering: Prepare data for modeling success

Filtering & Engineering

- Filter for the intended analysis population.
- Generate features & modify existing ones.

Feature Pre-selection

- Identify variables with clinical relevance.
- Remove features with extensive missingness.

Imputation

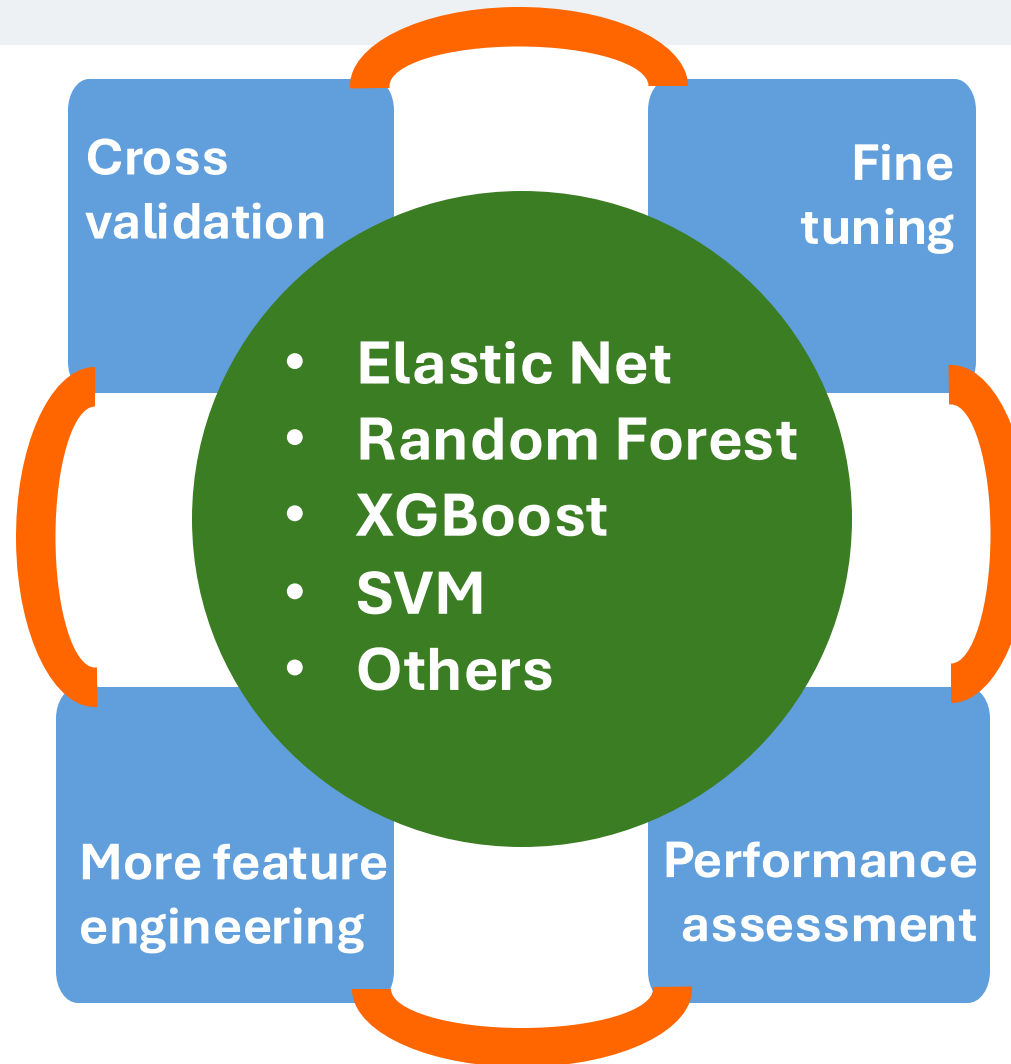
- Multiple imputation, KNN imputation, regression imputation etc.

Oversampling

- SMOTE, ADASYN etc.

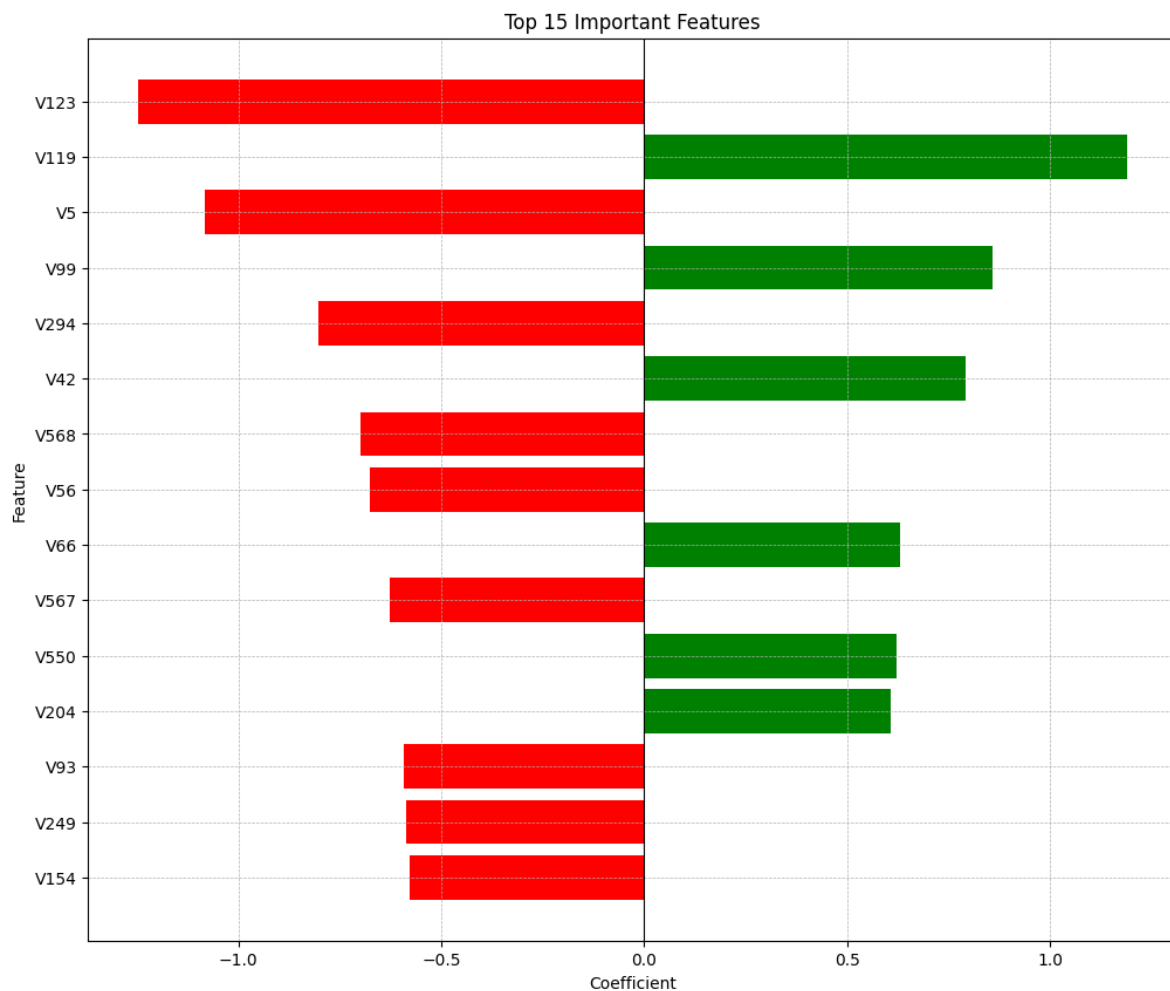
4

ML Modeling: Embrace the refinement loop



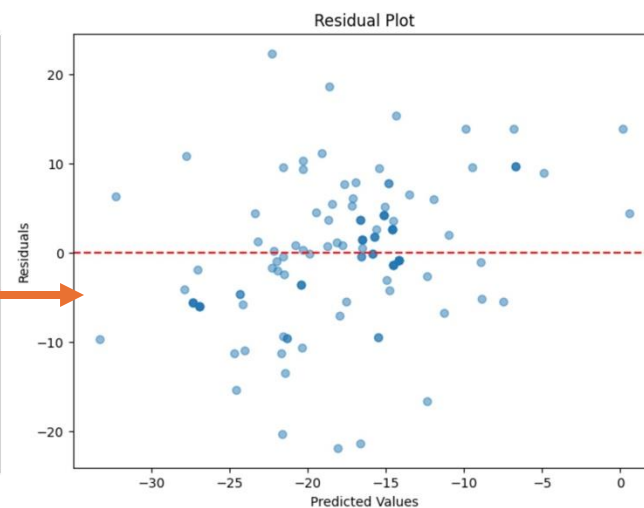
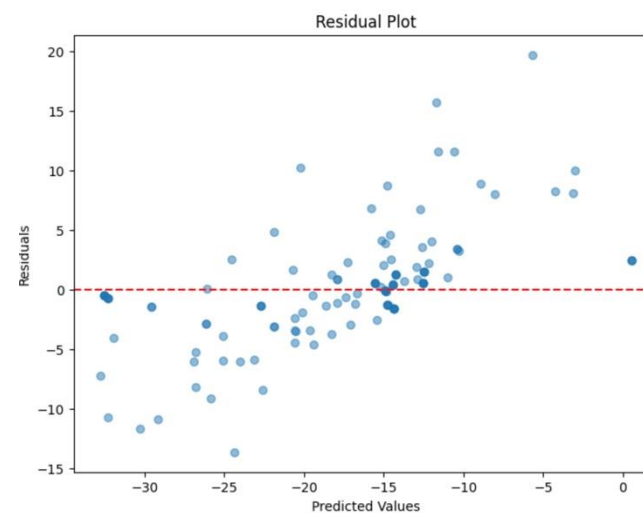
5

Reporting: Comprehensive, transparent, and graphically driven



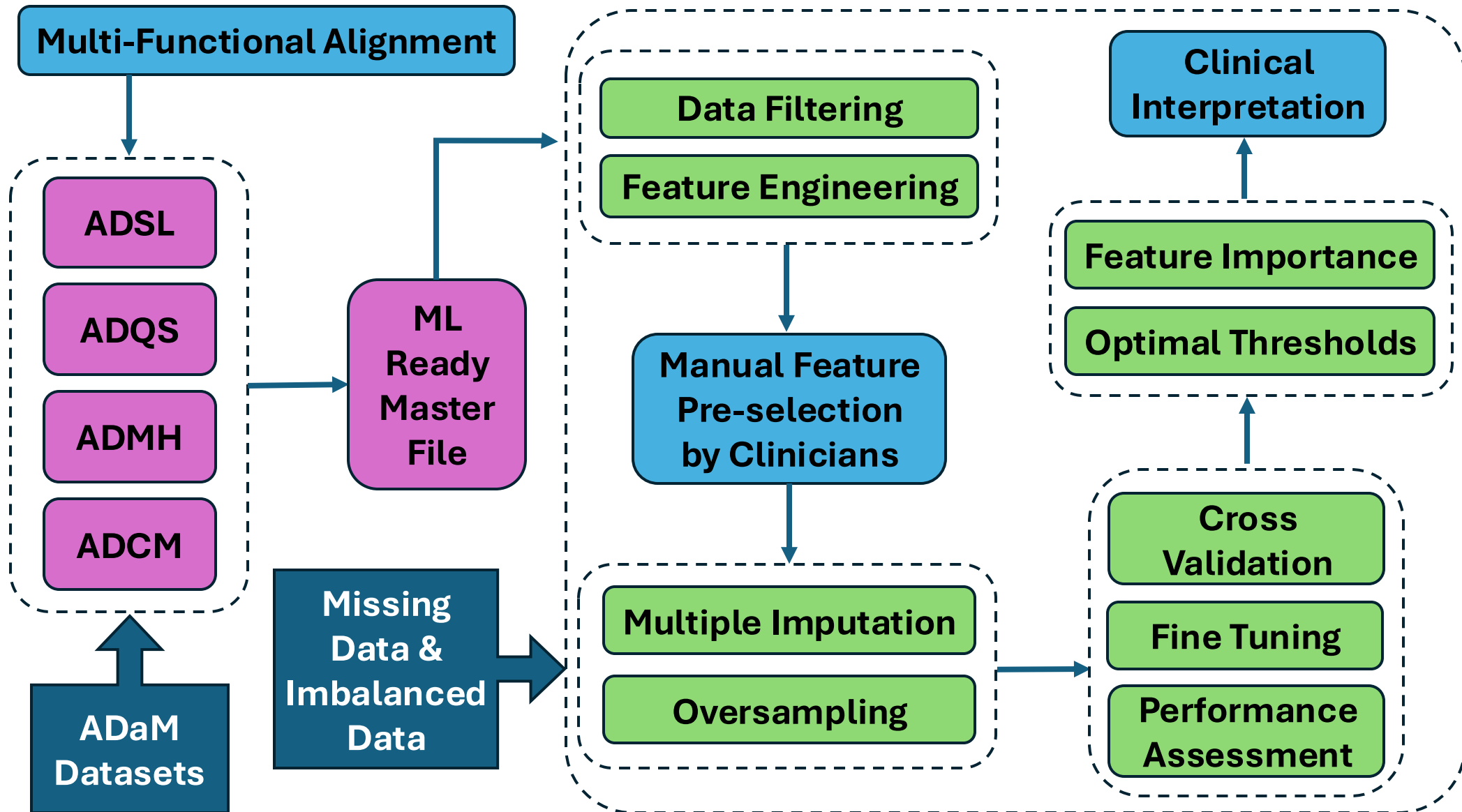
Model performance improvement after multiple imputation and oversampling

Metrics	Elastic Net	Random Forest
R ²	0.35 → 0.78	0.47 → 0.81
RMSE	10.77 → 7.83	8.99 → 4.62



Metric	Formular/ Confusion Matrix	Description
FPR	$FP/(FP + TN)$	False Positive Rate
FNR	$FN / (FN+TP)$	False Negative Rate
Sensitivity /Recall /TPR	$TP / (TP+FN)$	How many (%) of target people are correctly identified as target
Precision /PPV	$TP/(TP+FP)$	How many (%) of predicted target are actual target
Specificity /TNR	$TN/(TN+FP)$	How many (%) of control people are identified as control
Accuracy	$(TP + TN)/(TP+FP+FN+TN)$	How many (%) of total people correctly identified among total patients
F1-Score	$\frac{2 * Precision * Recall}{Precision + Recall}$	F1 takes both precision and recall into account to ultimately measure the accuracy
AUC	Area Under the Receiver Operating Characteristic (ROC) Curve	Validate the trade-off between Sensitivity and Specificity

Streamlined Workflow: A Practical Example



Future Directions



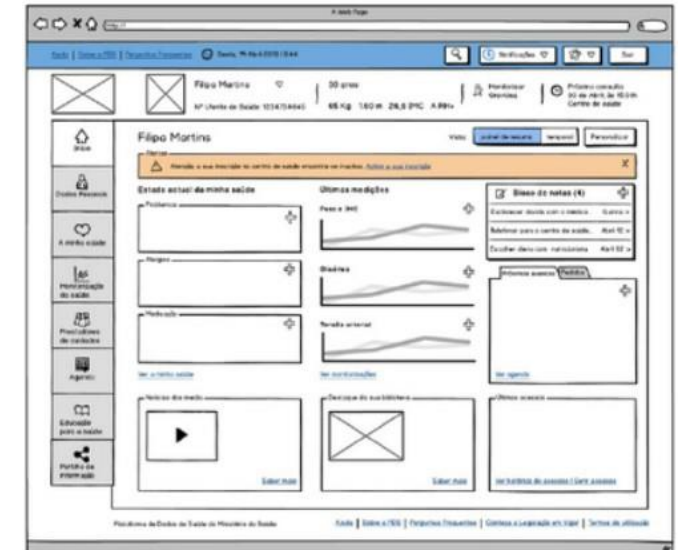
Predictive Assessment Tools: Forecast treatment response and dropout likelihood to improve trial design and decision-making.



Interactive Patient Profiling: Develop applications to help clinicians visualize participant profiles, including placebo risk and dropout probability for more informed decisions.



Real-Time Risk Assessment & Monitoring: Integrate ML-driven tools into ongoing clinical trials to enhance efficiency and optimize trial outcomes.



Acknowledgement and Contact Info

❖ Presenting and communication author:

Jing Dai (Jazz Pharma and Wayne State University)

Email: jdai@jazzpharma.com

❖ Co-authors:

Brian Shonesy (Jazz Pharma)

Selina (Jiayu) Hu (Northwestern University)

Ming Qiang Huang (former Jazz Pharma)