Predictable Delivery of Clinical Trials

MIT MBAn Students

Denis Sai

Edoardo Italia

MIT Faculty Advisor

Prof. Steve Graves

Takeda

Saurabh Awasthi

Melissa Chiasson-Umbehauer

Stephen Cue

Sneha Narayanan

Scion Li

Problem Statement

Illustrative Clinical Trial Pipeline

Planning → Recruitment → Treatment → **Selected segment** [Close Out → Analysis and Reporting]
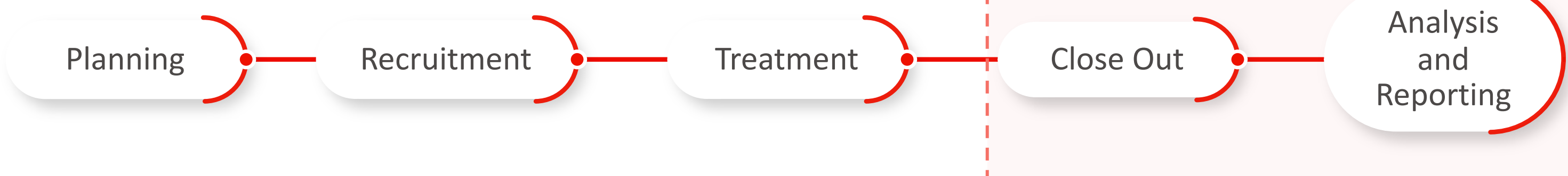
In navigating the complexity of clinical trials our **goal is to characterize** the performance of our selected clinical trial segment, as well as provide **pathways for improved operations.**

Impact

6% faster cycle times
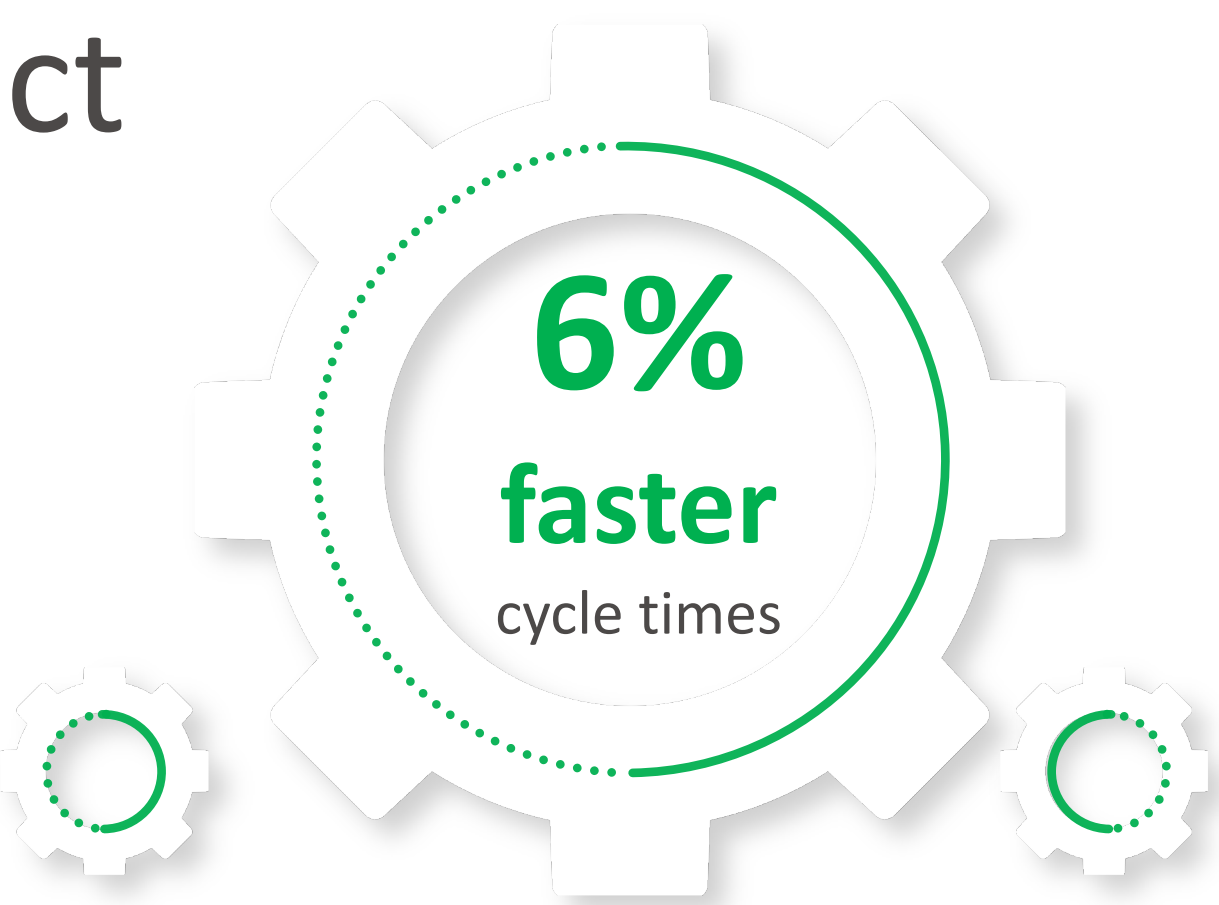
Upon leveraging numerous machine learning models, we quantified the impact of two statistically significant operational components. These were found to **affect both trial delivery and net present value (NPV)**.

Yearly Additional Net Present Value

$9 M Worst case

$39 M Conservative

$68 M Best case

Resulting from our machine learning insights, we can conduct a scenario analysis and **quantify the lift in NPV**. Upon a minimization of the identified key hurdles, we observe three potential lifts in NPV.

Deep Dive

Data
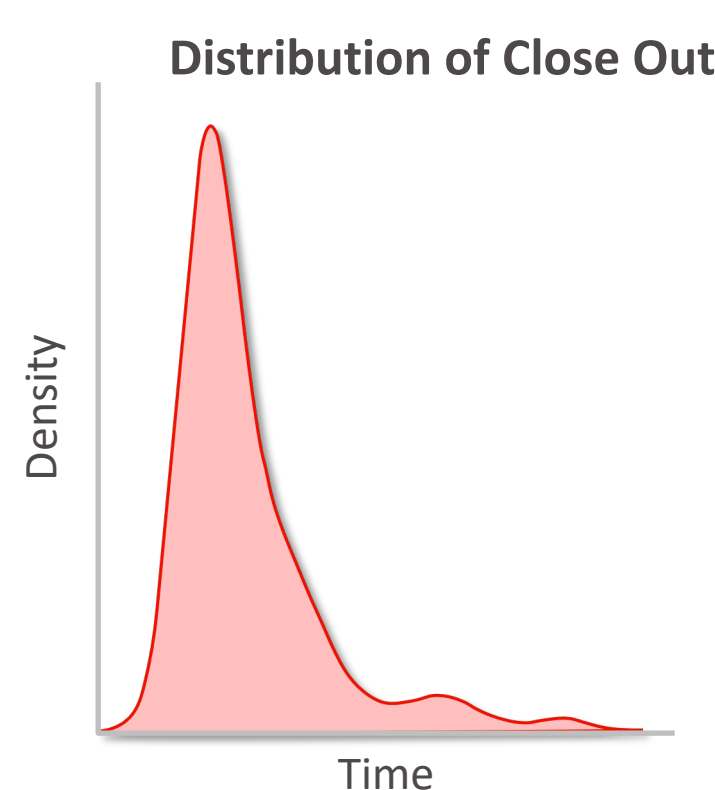
Few observations but numerous features to screen

Varying degrees of sparsity associated for each dataset

Complex data structures from a range of sources

Descriptive

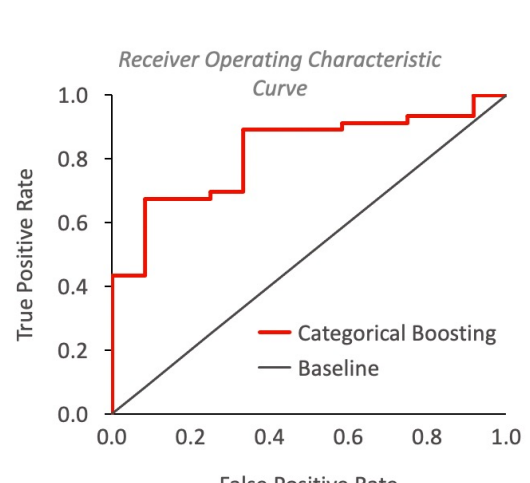Distribution of Close Out

Density vs Time

To model our dependent variable, **we leveraged several Machine Learning families of approaches:**

- Interpretable regressions
- Count and ensemble models
- Binary and multiclass classification

Predictive

We were able to **achieve strong Out-of-Sample performance** predicting **on-time delivery** of clinical trials aligned with internal time expectations.

We used a wide range of models, such as: Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, Categorical Boosting

Receiver Operating Characteristic Curve
True Positive Rate vs False Positive Rate
— Categorical Boosting
— Baseline

Out-of-Sample Area-Under Curve **0.82**

Time predictions **2** Statistically significant components

Trial delivery **3** Statistically significant parameters

Since we **care about interpretability** of our models, we found statistically significant parameters in two specification:
- Time predictions of specific trial segment
- On-Time delivery with the specified segment

Leveraging our results is going to be the next step in Takeda analysis

Takeda Pharmaceutical Company Limited