Good Site/Bad Site: An efficient strategy for clinical site allocation


Capstone Team
Maria Camila Marenco
Aziz Ayed

Faculty Advisor
Retsef Levi

Takeda Team
Saurabh Awasthi
Stephen Cue
Melissa Chiasson
Shujaullah Mohammed
Scion Li

1. Problem Statement and Motivation

Context: Without sites there would be no clinical trials and no new medications.

Takeda has **multiple clinical site options** to choose from and thus, **site performance is critical**

Overall project goal: Develop an analytics solution based on historical data to improve Takeda's site selection

Sub-objectives

1. Will a site be non-enrolling?
2. Will a site be a low-med or high enroller?
3. What's the time point when a site will never enroll?
4. What is the optimal site selection for a given study?

Why does it matter?

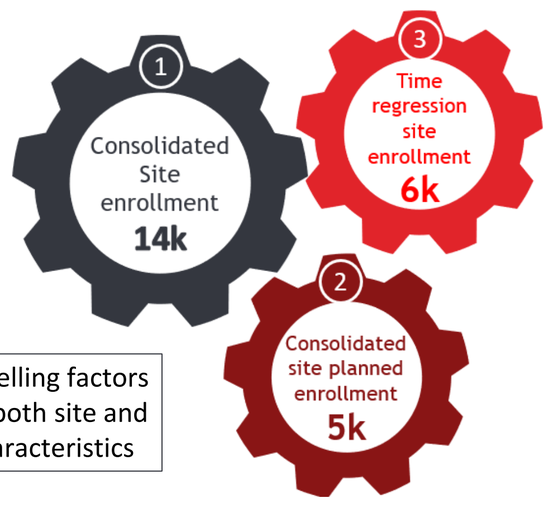
Getting drugs out faster to patients in need

Better allocate resources on new drugs development

Decrease costs and study delays:
- A typical trial can cost ~86M
- Delayed trials take +1-6 months

2. Data

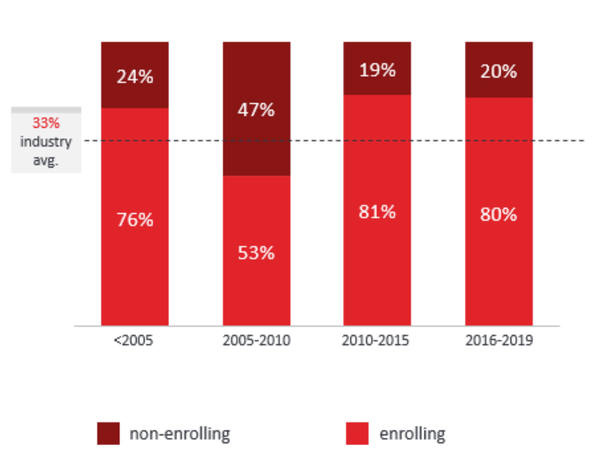
We created **3 distinct datasets** for our analysis leveraging internal trials data mostly after 2010



~140 modelling factors including both site and study characteristics

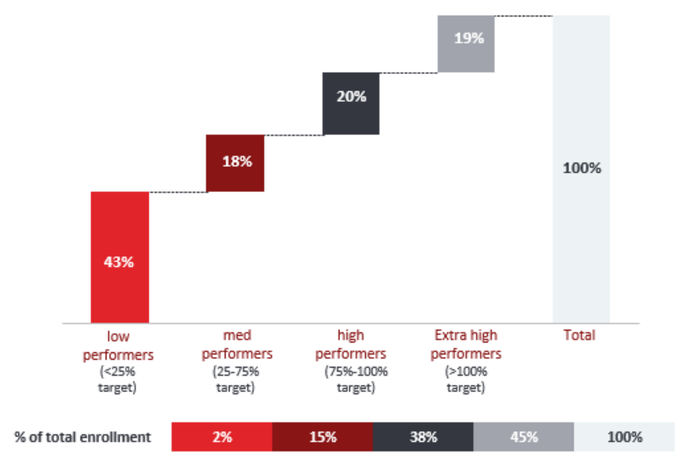
3. Exploratory Analysis

of enrolling vs non-enrolling studies at sites



~30% of all site-studies are **non enrolling**, with that % decreasing in recent years

# of sites per performance category



Higher performing sites contribute to most of total patient enrollment (~50%) while being the minority in number of sites (<20%)

4. Methodology

3 Predictive Machine Learning Models

1. **Classification** model to predict probability of non-enrolling sites
2. **Multi classification** model to predict low-med-high enrolling sites
3. **Survival model** (log-logistic AFT) to predict time inflection point of never enrolling

1 Prescriptive Optimization model

4. **Dynamic optimization model** for site selection:
   - Using **closed form expression** of classification models as constraints
   - Accounting for **complex interactions** with **dynamic optimization**
   - Maximize expected enrollment while minimizing costs
   - Control over **minimal proportion of high enroller** and maximal proportion of low enrollers
   - **Piecewise linear approximation** of the sigmoid

These 4 analytics models built will allow Takeda **to act on three different parts of the site selection & management** process

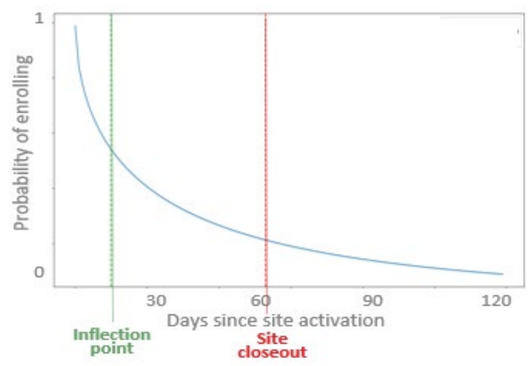
5. Results

Built **high performing** models:

0.93 AUC

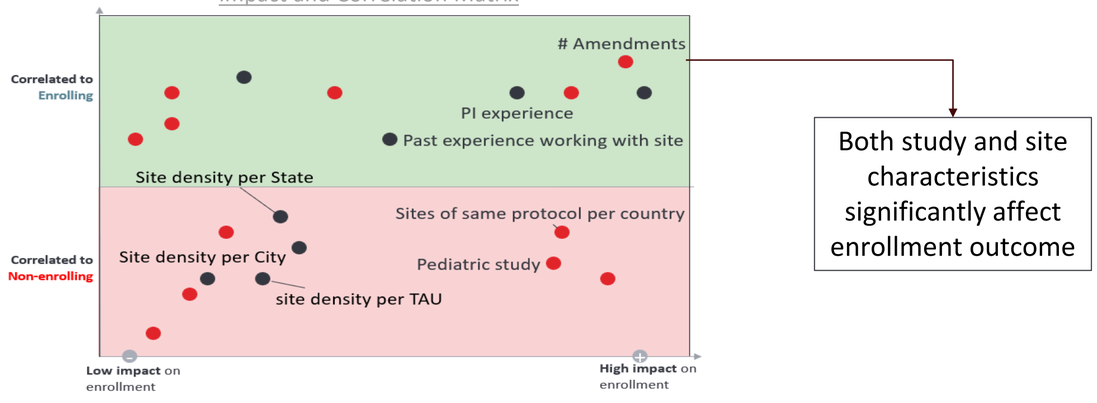
AUC for best performing model
2nd model:~0.8 AUC
3rd model: ~0.7 C-index

With personalized results on not only site selection but also site close out



That provided actionable recommendations Identified subset of most impactful site and study characteristics affecting enrollment

Impact and Correlation Matrix



6. Impact and Next Steps

Impact of our work

$200M

Avg. 5-year cost savings just by considering non-enrolling sites. Could even be more (i.e.: entry to market saving)

Accelerate drug development: Getting drugs out faster to patients – advancing society

Implementation

Currently testing our solution on a **two-study pilot**

Our project will be **implemented in the clinical analytics hub** in a 2–3-year horizon

Future Areas of work

Data collecting: Incorporate external data and plan to collect further information from CROs

Expand scope: Include in analysis other KPIs (I.e.: retention) and to include ongoing effect of other actions

Takeda Pharmaceutical Company Limited