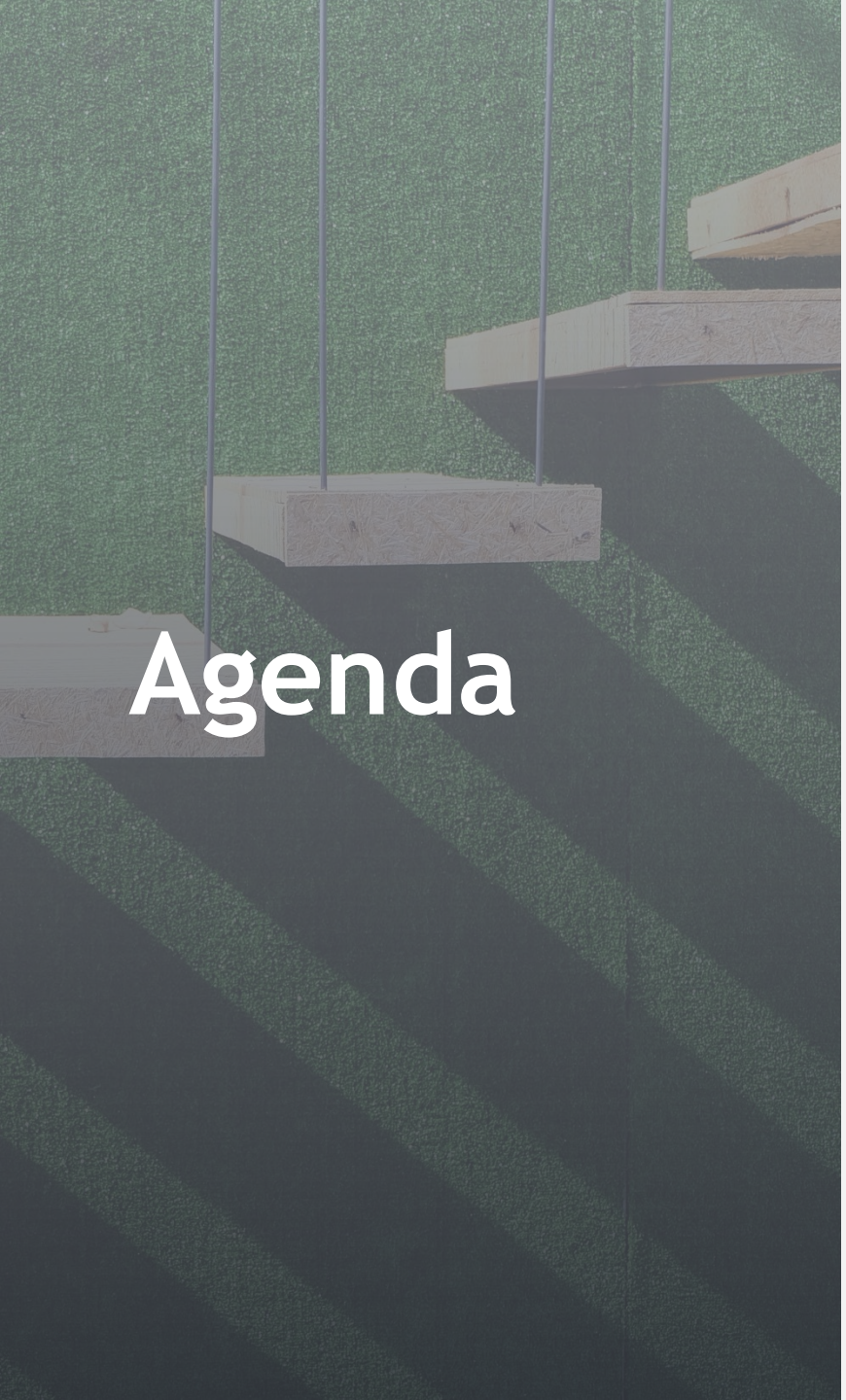
Good Site/Bad Site: An Efficient Strategy for Clinical Site Allocation

August 19, 2022

Maria Camila Marenco & Aziz Ayed

Better Health, Brighter Future



Agenda

1. Context and Problem Statement
2. Methodology
3. Results and Impact
4. Next Steps and Future Work

Clinical trials are a pivotal aspect of Takeda's business

Takeda is a patient-focused, values-based, **R&D driven** global biopharmaceutical company

Clinical trials are pivotal in drug development to assess the safety, effectiveness and efficacy of a drug to receive approval

A drug it must go through several phases of research before it's public and **without clinical trials there would be no new medications**

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

Currently, Takeda is exploring site performance opportunities in 2 dimensions

Non-enrollment

- **Context:** Sites that have enrolled 0 patients for a finished study
- **Improvement opportunity:** Significantly high levels of non-enrollment as well as, non-enrolling sites staying open for long periods of time

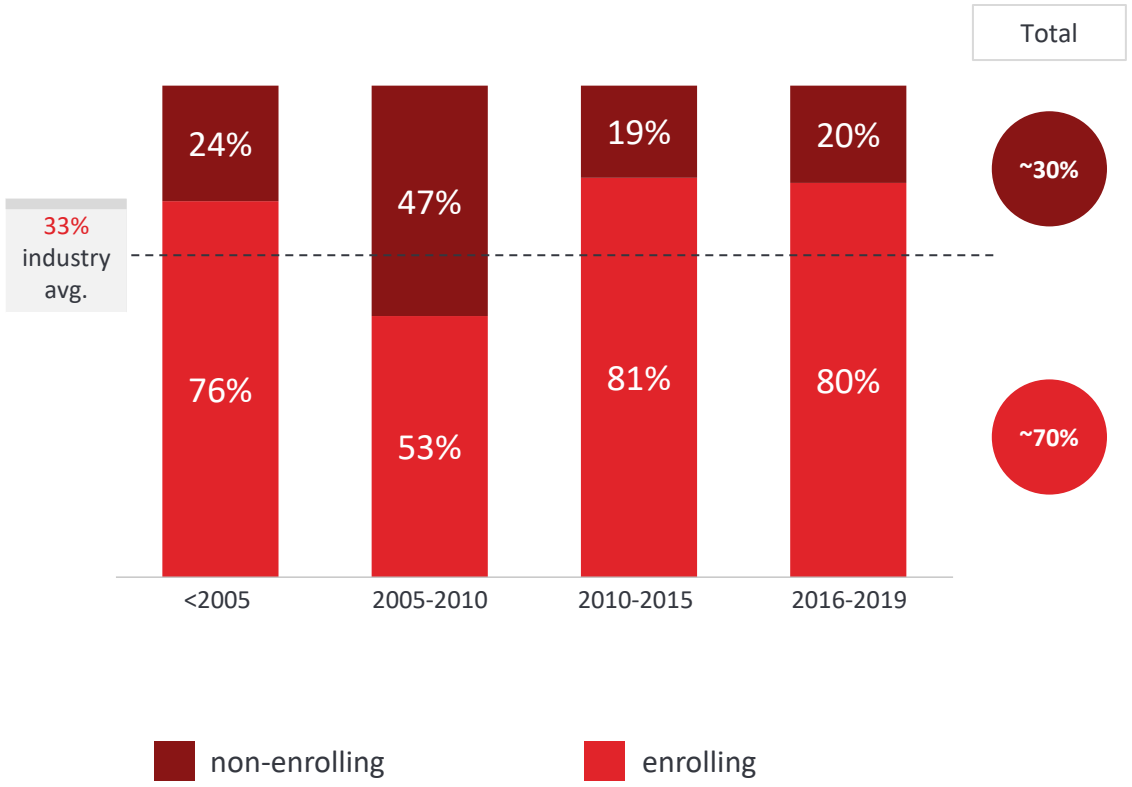
Not meeting target

- **Context**: Sites that have enrolled less patients than their planned target
- **Improvement opportunity** : Minority of sites (high performers) contribute to most of total enrollment

After exploring and analyzing historical clinical trials data, we confirmed opportunities around non-enrollment and not meeting target enrollment

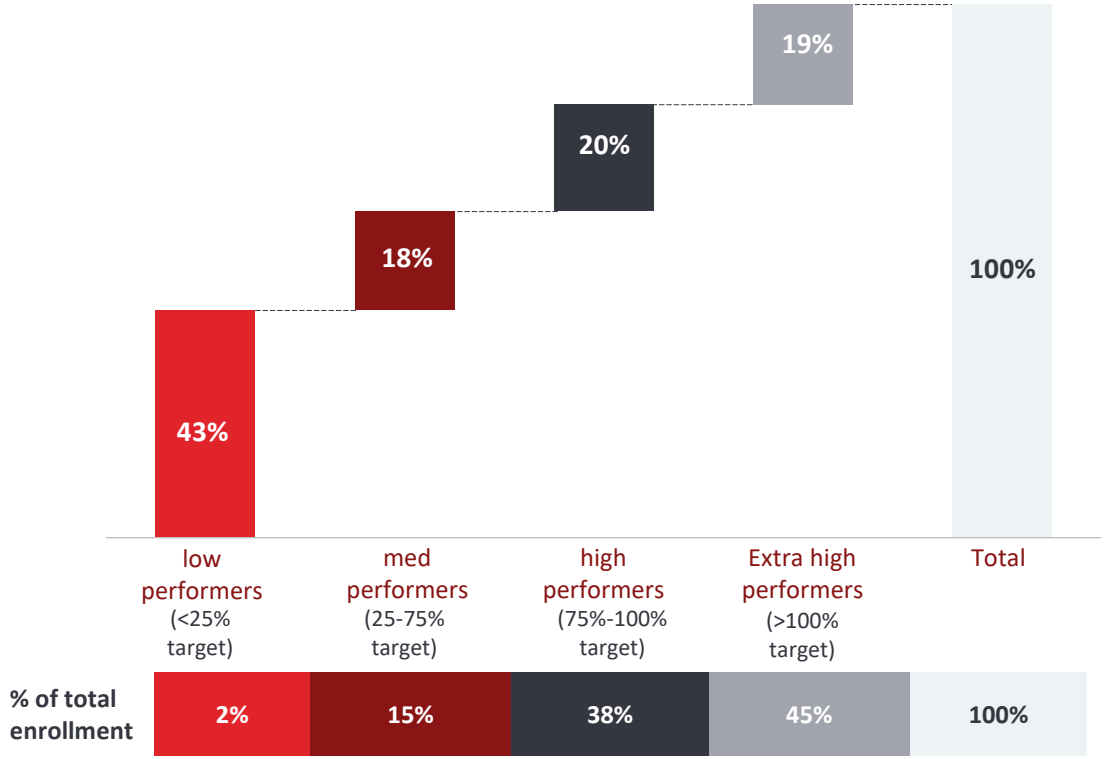
~30% of all historical site-studies are **non-enrolling,** with that % decreasing in recent years but still significant

# of enrolling vs non-enrolling studies at sites



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

Number of sites per performance category



Given this context, our project’s goal is to use analytics on historical data to improve Takeda’s site selection and management

Our project aims to utilize analytics to answer 4 key business questions around enrollment performance and site selection

3 questions are to be answered in the 1st phase of the project (predictive phase) and 1st in the 2nd (prescriptive phase)

Phase 1 (Predictive)

Phase 2 (prescriptive)

1

What's the probability a site will be **non-enrolling?**

Including sub-questions as:

- What **attributes** best characterize non-enrolling?
- Do the **performance hypothesis** identified affect enrollment according to historical data?

2

What's the probability a site be a **low-medium or high enroller?**

Including sub-questions as:

- What **attributes** best characterize performance?
- Do the **performance hypothesis** identified affect enrollment according to historical data?

3

What's the **tipping time point** when a site is likely to **never enroll a patient** from that moment on?

Including sub-questions as:

- How does this time-inflection point **vary between attributes** (i.e.: Therapeutic Area, phase, etc.)?

4

What's the **optimal site selection** for a given study?

Including sub-questions as:

- Optimal assignment of sites for a given study to maximize enrollment and minimize costs

Note: The Site health characterization along various dimensions is subject to further data availability and is currently P2

Methodology: To answer these four key questions we built 4 Machine Learning models

3 questions are to be answered in the 1st phase of the project (predictive phase) and 1st in the 2nd (prescriptive phase)

Phase 1 (Predictive)

Phase 2 (prescriptive)

1

What's the probability a site will be **non-enrolling?**

Model 1:
Machine Learning
classification model
- ~140 factors
- 14.5k observations

2

What's the probability a site be a **low-medium or high enroller?**

Model 2:
Machine Learning
multi- classification model
- ~140 factors
- 5k observations

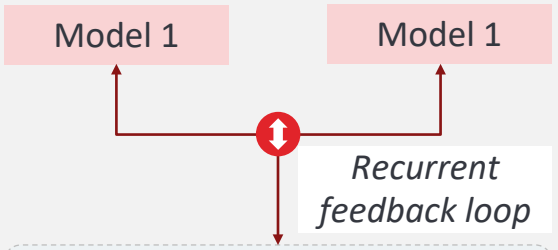
3

What's the **tipping time point** when a site is likely to **never enroll a patient** from that moment on?

Model 3:
Survival model
log-logistic AFT
- ~140 factors
- 6k observations

4

What's the **optimal site selection** for a given study?



Model 4:
Dynamic Mixed integer Optimization model
Feeds from model 1 and 2 to take into account complex site interactions

Note: The Site health characterization along various dimensions is subject to further data availability and is currently P2

Detail: The dynamic optimization model for site selection is currently tested in a two study pilot

Model objective

Maximize enrollment while **minimizing costs**

Using closed form expression of class. models as constraints

Constraints

Accounting for **complex interactions** with dynamic optimization:

- Control over **minimal proportion of high enroller** and max proportion of low enrollers
- **Piecewise linear approximation** of sigmoid
- **Geographical restrictions** and density constraints

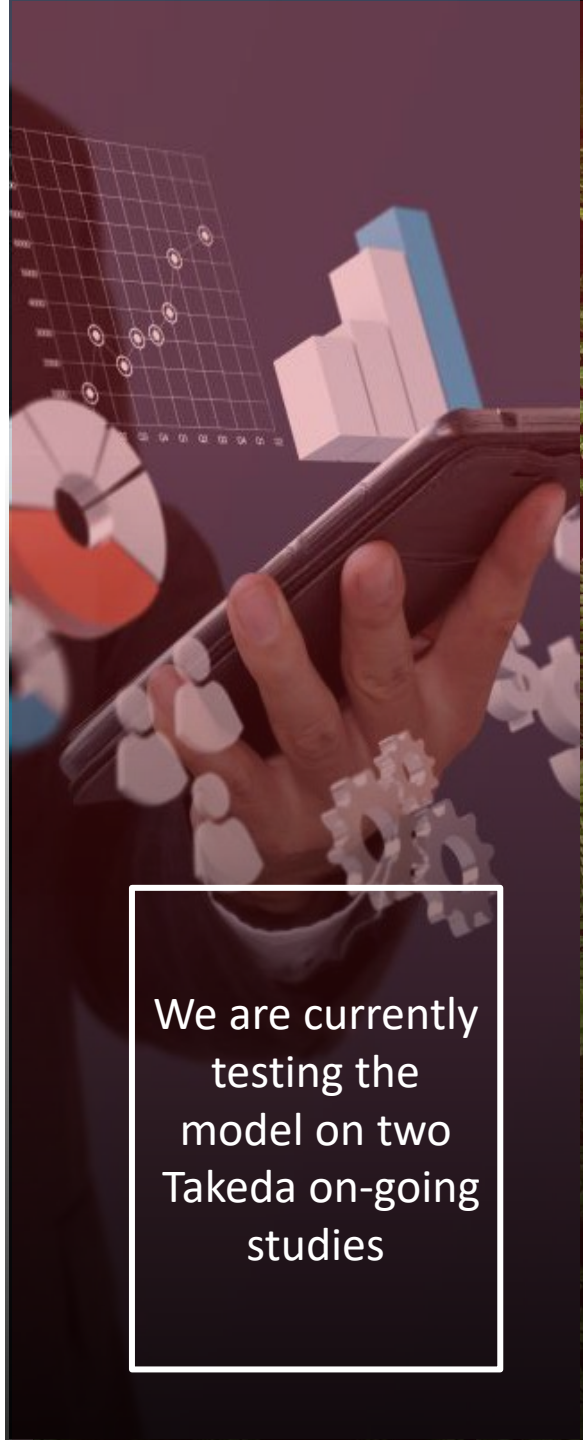
Input data

Both **study and site** characteristics:

- **Study characteristics** previously discussed
- **Pool of potential sites** to select from and characteristics previously discussed

Model output

Which sites to select for that given study, including the optimal number of sites to select



Results: Built high performing models, with personalized results that provided actionable recommendations for Takeda's site-studies

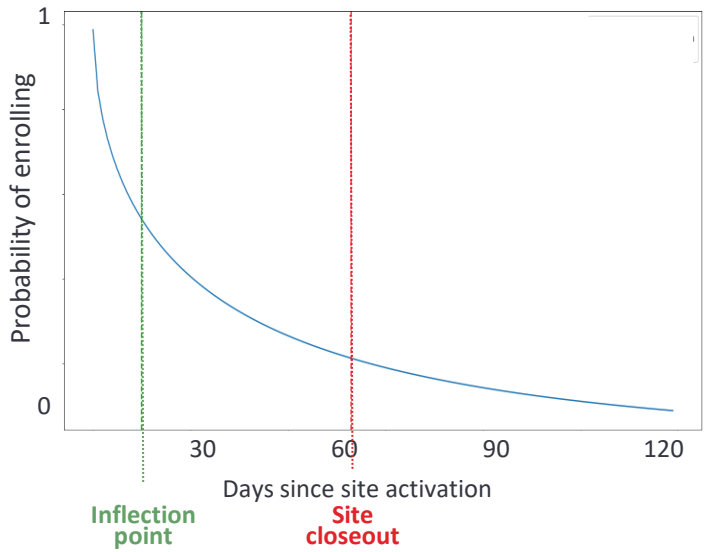
Built high performing models for to improve site selection..

Best performing model (XGBoost), allows Takeda to identify non-enrolling sites with very high confidence!

- With +33% increase vs. current site selection operations
- 2nd model had a ~0.8 AUC and 3rd model a ~0.7 C-index
- Dynamic optimization results account for complex interactions of chosen sites (i.e., geography)

With personalized results for distinct site-studies..

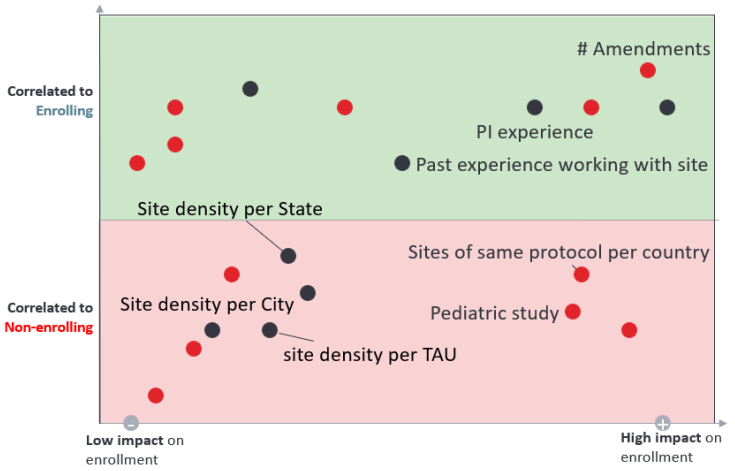
Survival curve example



Takeda not only has information on what sites to select but also **when to close a specific site out**

That provided actionable recommendations for Takeda

Matrix: Impact and correlation of Site and Study characteristics with enrollment



Identified most impactful site and study characteristics affecting enrollment **that Takeda can act on**

- Based on SHAPLEY and tree analysis of best performing model

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

Site Selection and Management Process

1 **Model 1:** First filter and narrow down potential site options, by **discarding sites with a high probability of being non-enrolling**

2 **Model 2:** Second filter to **target highest performing** sites for a given study

4 **Model 4:** Using as an input model 1 and 2, **recommend optimal site selection** for a given study and a given pool of sites to choose from

3 **Model 3:** For further protection, Takeda can identify the moment in time when a site has **high probability of never enrolling a patient** and **act earlier to move to site close-out**

Answering the previous questions can significantly impact Takeda on 2 key dimensions

Impacting patients and society:

- **Accelerate clinical studies** to get **drugs out faster to patients** that need them the most
- **Better allocate resources** on new drug developments for society

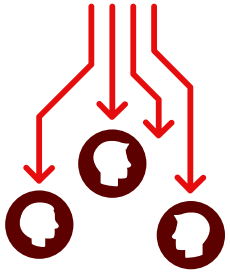
Impacting costs and study delays:

- **~200M USD cost savings** in 5-year time frame only by non-enrolling site reduction
- **Minimizing study delays** which are currently estimated around 1-6 months per study

Next Steps and Future Work

Future work

- **Data collecting**: Incorporate external data and devise plan to collect further information from CROs
- **Expand scope**: Include in analysis other performance metrics (I.e.: retention, screening) and to include effect of current actions on enrollment



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

CROs: contract research organizations

Thank you!

Better Health, Brighter Future