

TFS mentors: John Van Hoy, Yash Shah, Heather Bryson, Debra Shaumberg

MIT Faculty Advisor: James Butler

ClinOpt

Clinical Optimization for Smarter,
Low-Deviation Trial Design



Noe Bertramo



Victor Zhuang

PROJECT OVERVIEW

Problem Statement

90%
of drug
candidates fail in
clinical trials
→ Nearly half due
to protocol
design &
execution flaws



ClinOpt empowers
clinical scientists to
design smarter, low-
deviation protocols.



Uses **historical trial data** to assess how
design choices like
eligibility or visit
schedules **affect**
deviations.



By turning intuition
into evidence,
ClinOpt helps
reduce trial
complexity and
improve **execution**
success.

Data



1000+ Protocols

- Covers 23+ specialties
- Free-text data format
- Rich data (trial endpoints, patient eligibility criteria...)



1.4M+ Deviations

- Sourced from Thermo Fisher internal deviation logs
- Categorized by types
- Text descriptions and status

PROJECT GOAL

Translate clinical intuition into data-driven protocol design decisions

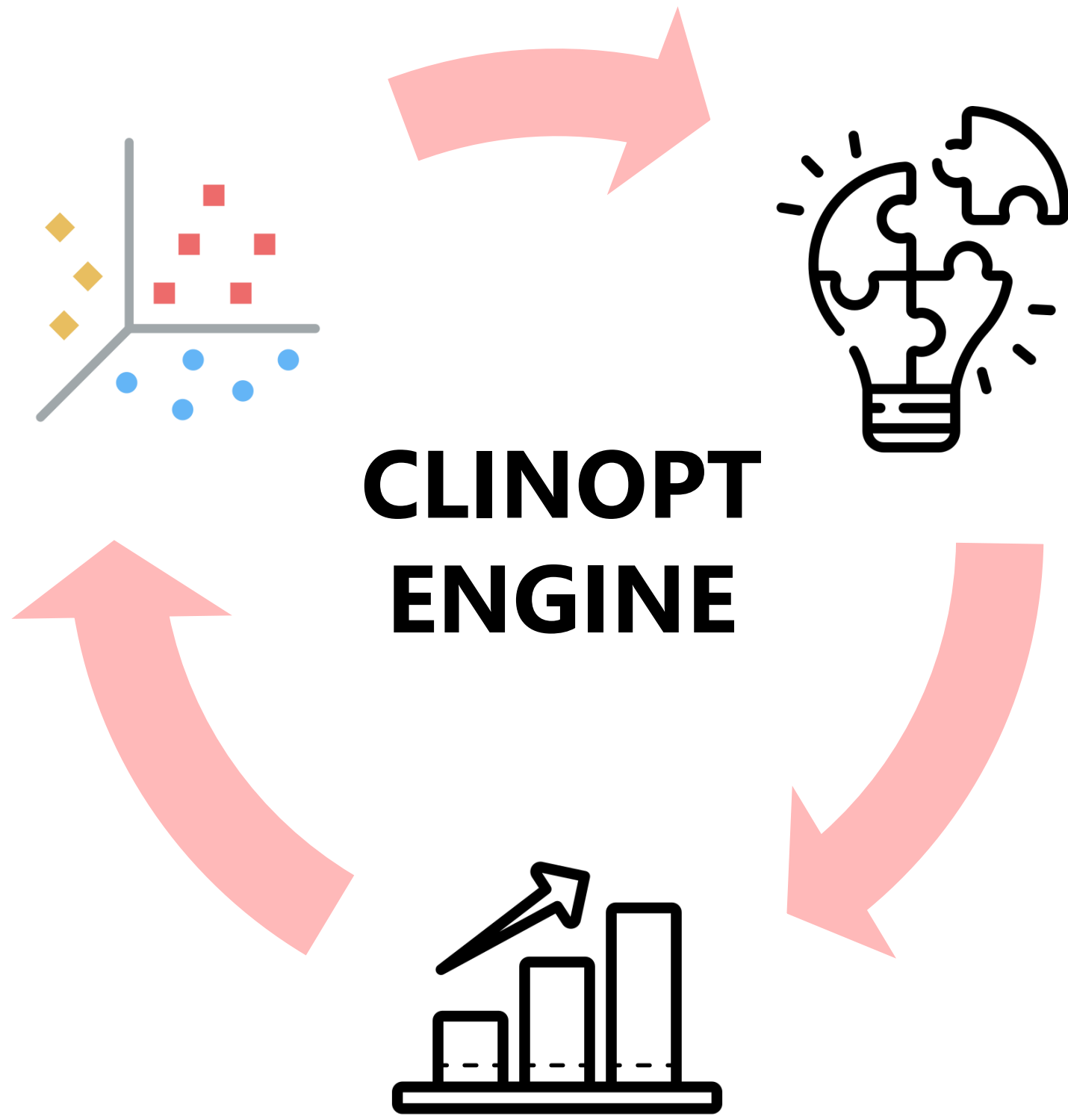
METHODOLOGY

1. Smart Clustering

Goal: Establish baseline deviation rates

- Group** trials by features (e.g. therapeutic area)
- Selection** based on mutual information, χ^2 and clinical relevance.
- Ensures comparisons are made across **meaningful subgroups**

Features	Mutual Information
Sponsor	4.13
Therapeutic Area	2.32
Phase	1.31



2. Feature Generation

Goal: Let scientists statistically test design hypothesis

- Users frame protocol-based **questions**
- NLP agent **extract and verify structured binary/categorical information**
- Supports criteria like “Is having a history of strokes an exclusion criteria?”
- Features are meant to **explain deviation count variation** within each subgroup

3. Causal Inference (DML)

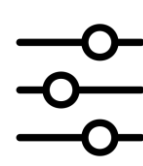
Double Machine Learning (DML)

- A technique that separates prediction from inference
- Controls confounders to isolate causal effects



Setup

- Grouped trials (e.g. ophthalmology)
- Outcome: deviation count
- Features: initial protocol parameters and user-defined treatments



Adjustments

- Confounders: trial phase, sponsor, country...
- Use ML to estimate nuisance
- Balance covariates

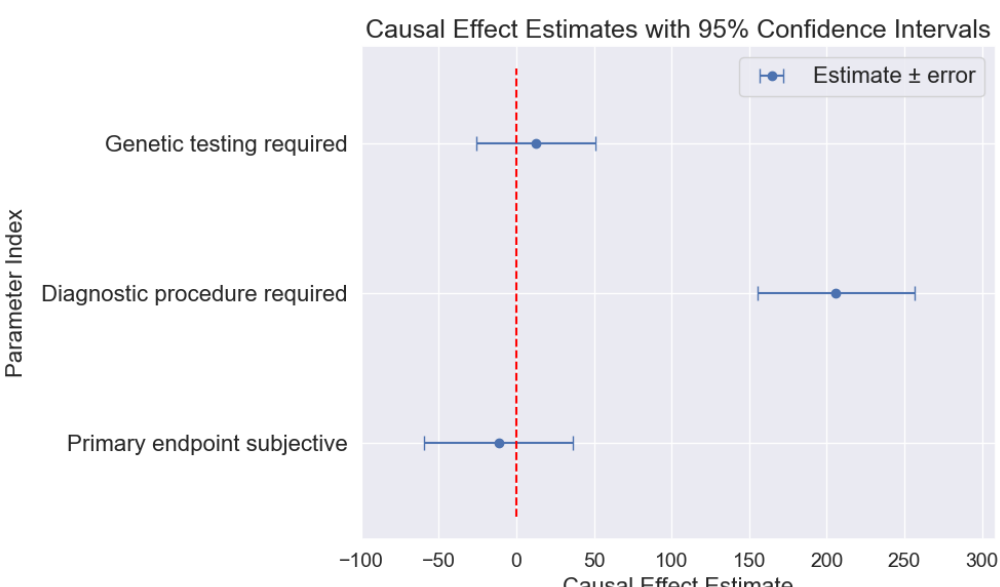


Results

- Estimate causal effect size
- Standard Error, Confidence Intervals
- Feature importance ranking

RESULTS

Study Case



- Tested ClinOpt on oncology trials to assess which features drive deviation rates.
- Found that inclusion criteria requiring diagnostic procedures (e.g. MRI, CT...) **significantly** increase deviations.

Business Impact

50+



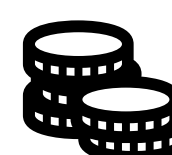
Potential
hours saved
per case

↓ 70%



Manual
work

\$37.5k –
\$119k



Potential
savings per
case