Sharing is Caring:
Investigation Load Balancing

Pfizer Team:
Abby Garrett & Jonathan Lowe

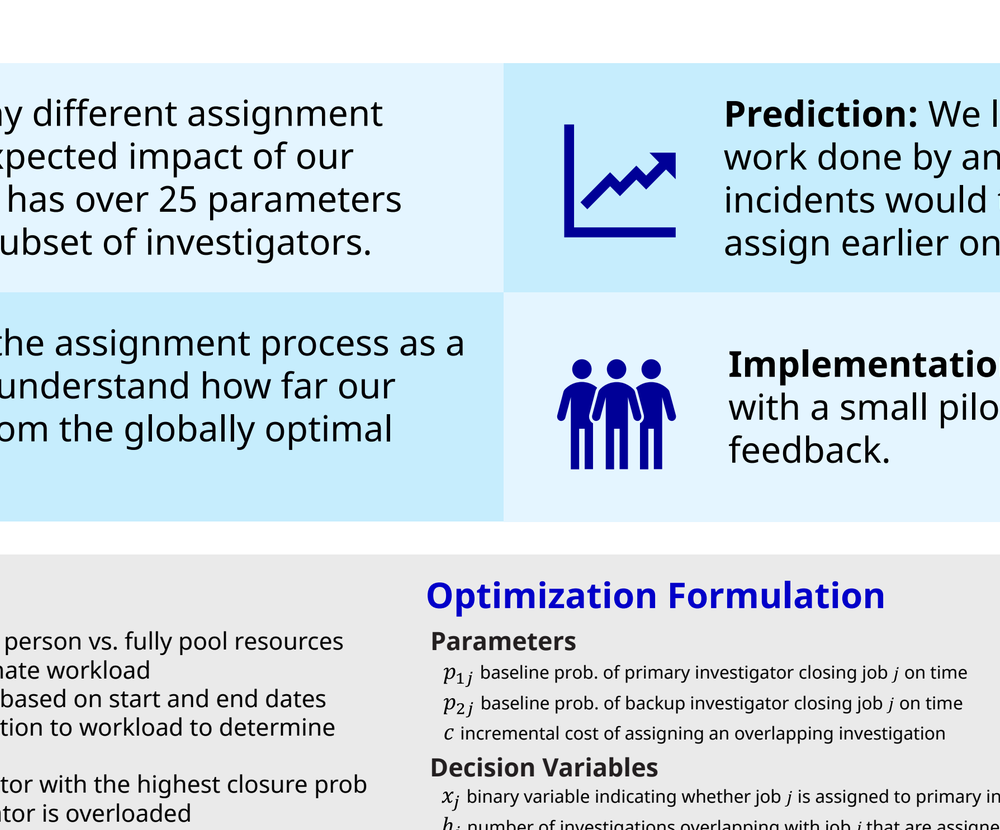
Faculty Advisor:
Prof. Daniel Freund

Capstone Team:
Stephanie Franklin & Shaun Gan

1 Problem Statement

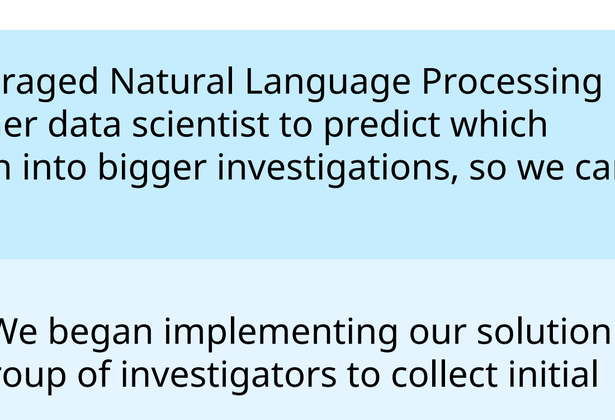
Overall project goal: Develop an analytics solution that can help Pfizer improve on-time closure of investigations.

What is an investigation?
Occasionally, something unexpected happens at the site: say a roof leak, or foreign material in the product - and investigators have to determine what went wrong and isloate the root cause. These investigations are supposed to take 30 days but investigators often have trouble meeting this deadline.



A big part of our project was finding an impactful area of focus. Can you think of any other hypotheses to explore?

2 Exploratory Data Analysis



Our area of focus: After extensive exploratory data analysis and conversations with multiple sites, we settled on load balancing as the biggest opportunity.

Key motivating findings: With each new overlapping investigation an investigator has, their likelihood of on-time closure decreases by ~3%. However, not everyone gets overloaded at the same time, so there is opportunity to load balance.

3 Methodology

Simulation: We simulated many different assignment strategies to understand the expected impact of our proposed logic. Our simulation has over 25 parameters and can be run for any site or subset of investigators.

Prediction: We leveraged Natural Language Processing work done by another data scientist to predict which incidents would turn into bigger investigations, so we can assign earlier on.

Optimization: We formulated the assignment process as a linear optimization problem to understand how far our online assignment process is from the globally optimal solution.

Implementation: We began implementing our solution with a small pilot group of investigators to collect initial feedback.

Example Simulation Parameters

- **Full pooling vs. one backup**: one specific backup for each person vs. fully pool resources
- **Overlap vs. ongoing**: determines metric used to approximate workload
- **Investigator dropouts**: investigators add in and drop out based on start and end dates
- **Assign via closure prob**: use investigator skill level in addition to workload to determine assignments, or just load balance
- **Maximize closure prob**: proactively seek out the investigator with the highest closure prob
- **Capacity threshold**: value that determines when investigator is overloaded
- **Backup penalty**: penalty to assume backup has less specialized knowledge
- **Investigation types**: which type of investigations to include
- **Minimum investigator experience**: min. # of investigations an investigator must have done to be eligible for assignment
- **Site, business unit, area manager, etc.**: run simulation for subset of site/investigators
- **Inactivity period**: assumption, how many days before we assume someone is inactive
- **Starting date(s)**: date range(s) to use for capacity assumptions and for simulation

Optimization Formulation

Parameters

p_{1j} baseline prob. of primary investigator closing job j on time
p_{2j} baseline prob. of backup investigator closing job j on time
c incremental cost of assigning an overlapping investigation

Decision Variables

x_j binary variable indicating whether job j is assigned to primary investigator or backup
h_j number of investigations overlapping with job j that are assigned to the same investigator

Sets

$j \in J$ set of all jobs (investigations)
$P_j \in J$ set of all jobs k that overlap with job j based on j and k's primary investigators
$B_j \in J$ set of all jobs k that overlap with job j based on j and k's backup investigators
$S_j \in J$ set of all jobs k that overlap with job j based on j's primary investigator and k's backup investigator
$T_j \in J$ set of all jobs k that overlap with job j based on j's backup investigator and k's primary investigator

Objective Function

$$\max \sum_j p_{1j}x_j + (1-x_j)p_{2j} - c * h_j$$

Constraints

$$h_j \geq \sum_{k \in P_j} x_k + \sum_{k \in S_j} (1-x_k) - M(1-x_j) \quad \forall j$$

$$h_j \geq \sum_{k \in T_j} x_k + \sum_{k \in B_j} (1-x_k) - M x_j \quad \forall j$$

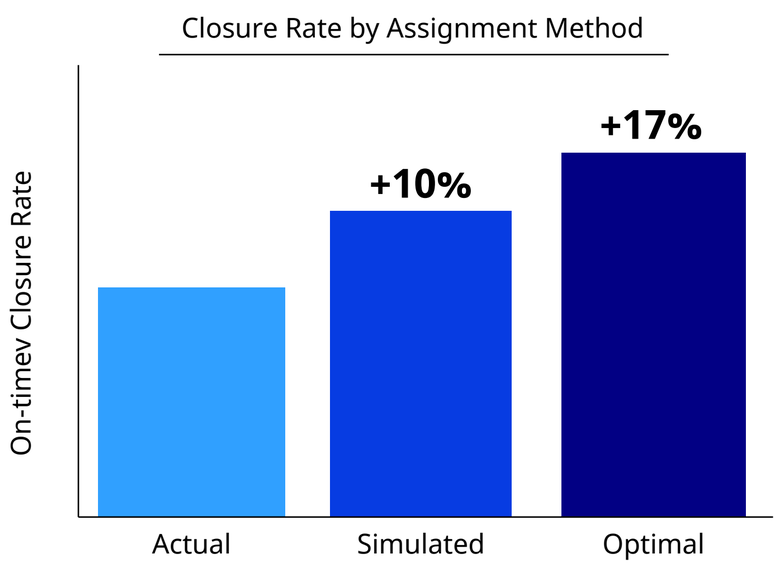
$$x_j \in \{0,1\}$$

$$h_j \geq 0 \quad \forall j$$

4 Results and Discussion

Online assignment process **closes >50% of gap** to offline optimal solution. This is a strong result, as real-time logic can never match a solution that has oracle knowledge of future investigations.

On-time closure uplift is strongest with broader resource pooling and more complex assignment logic, with **up to 13% uplift** projected across focus business unit.



> *I wish I had this all along. I'm happy to go along with your recommendations.*
>
> - Investigator on the pilot team

Breakthroughs that change patients' lives ®