



CLINICALAI

Providing data driven estimations for
clinical trial durations

Adeline Chin, Cynthia Xu, Jooyeon Hahm

2024 Summer
210 Capstone
Section 6

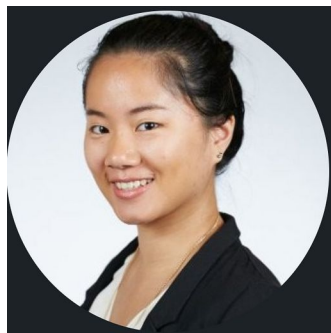


Our Team



ADELINE CHIN

Subject matter
expert
Frontend Developer



CYNTHIA XU

Data & ML
Engineering Lead
Backend Developer



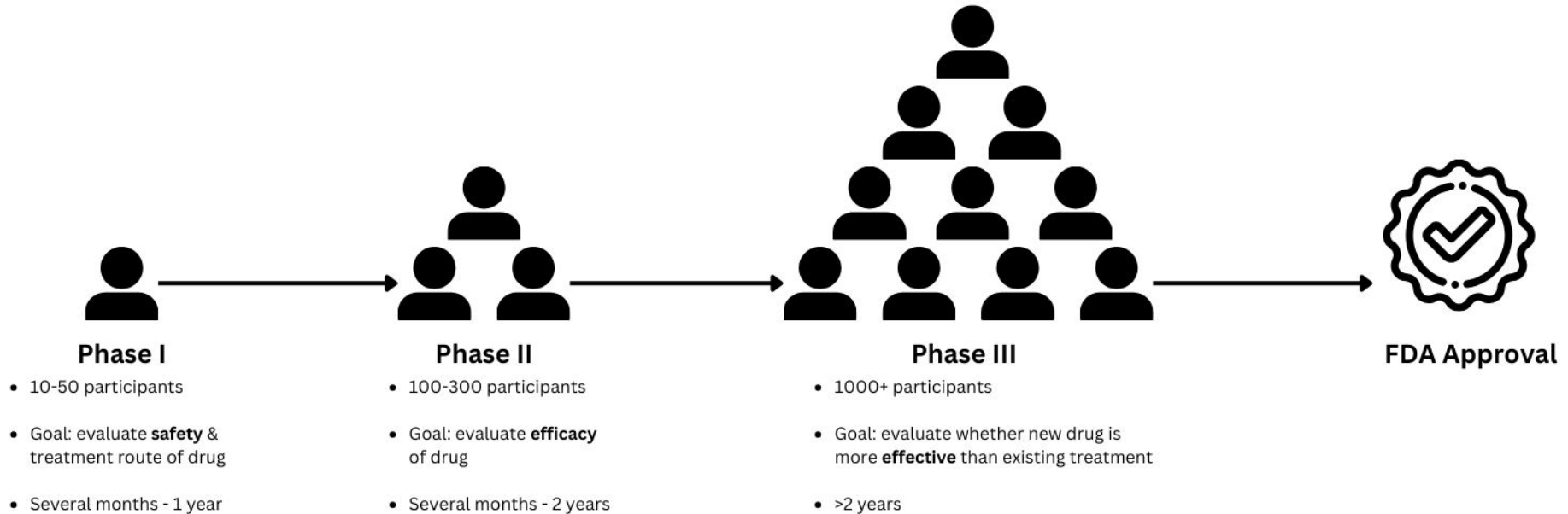
JOOYEON HAHM

ML Research & DL
Lead
Website Developer

Motivation



Clinical Trials are essential for the innovation and delivery of new drugs



Phase 3 clinical trials have a high risk of failure due to complexity and duration

BIOTECH

UPDATE: Acelyrin points to CRO error that could explain shocking phase 3 failure

By Annalee Armstrong · Nov 28, 2023 5:00am

INDUSTRY NEWS

¹Armstrong, A. (2023, November 28). Update: Acelyrin points to Cro error that could explain shocking phase 3 failure. Fierce Biotech.

²Schmidt, H. (2023, August 11). Nektar sues Eli Lilly for incorrect clinical trial results. PharmaNewsIntelligence.

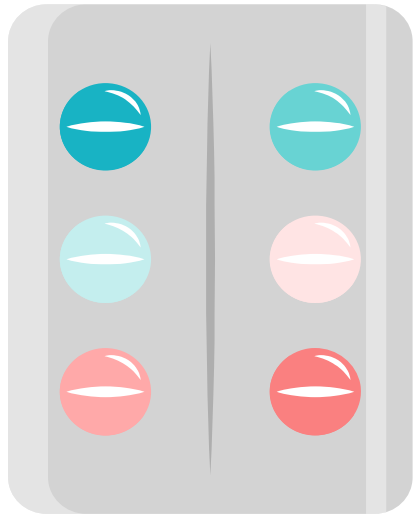
CLINICAL

Nektar Sues Eli Lilly for Incorrect Clinical Trial Results

The immunology firm is suing its big pharma partner after clinical trial data failed to fully demonstrate the positive impact of an atopic dermatitis drug.

Aug 11, 2023

Despite a large market size, a publicly accessible tool for predicting trial durations is lacking.




Estimated **\$55.86 billion** global market value for CROs

39,722 new clinical trials registered in 2023 at **clinicaltrials.gov**

1. Pawar, N. (2024). Clinical Trials Market Size, Growth and Statistics 2030. Vision Research Reports. <https://www.visionresearchreports.com/report/checkout/38176>
2. National Library of Medicine: National Center for Biotechnology Information. (2024). Trends and Charts on Registered Studies. National Library of Medicine: National Center for Biotechnology Information <https://clinicaltrials.gov/about-site/trends-charts>

ClinicalTrials.gov is a global registry for clinical trials. We leveraged 19K post-2011 oncology studies to build our models.

 An official website of the United States government [Here's how you know](#) ▾

 National Library of Medicine
National Center for Biotechnology Information

PRS Login

ClinicalTrials.gov

[Find Studies](#) ▾ [Study Basics](#) ▾ [Submit Studies](#) ▾ [Data and API](#) ▾ [Policy](#) ▾ [About](#) ▾

[My Saved Studies \(0\)](#) →

ClinicalTrials.gov is a place to learn about clinical studies from around the world.



The U.S. government does not review or approve the safety and science of all studies listed on this website.

+

Read our full [disclaimer](#) for details.

CLINICAL

Existing trial management solutions fail to specifically predict trial durations.

Non-ML solutions



ML solutions



Our goal

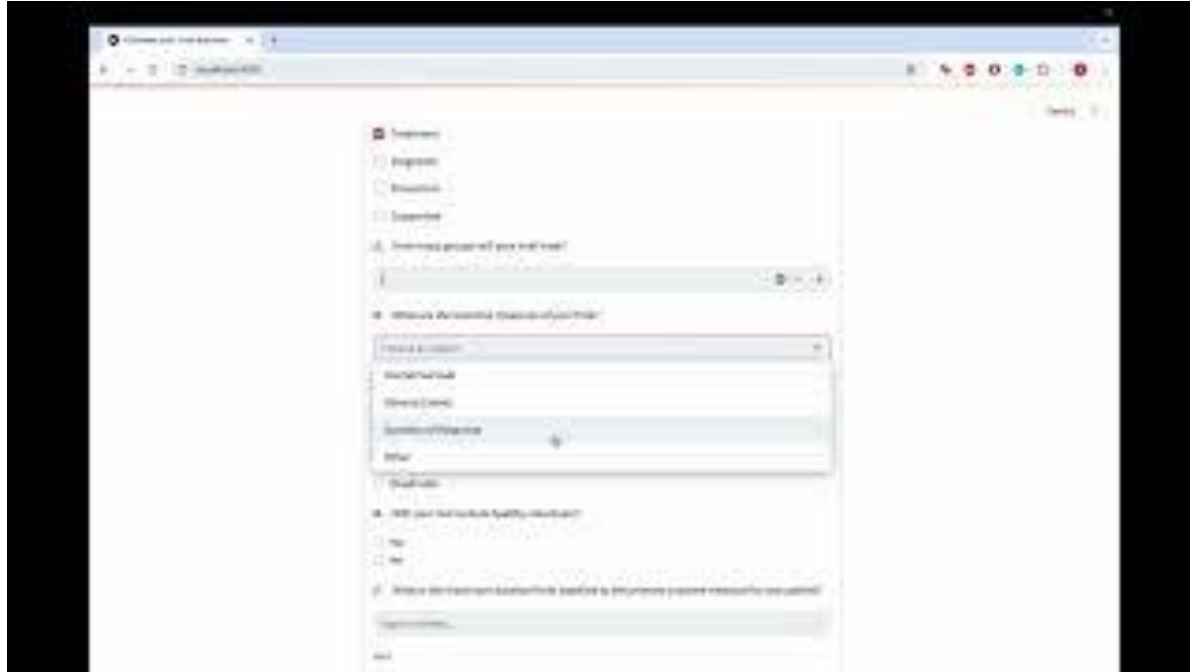
Develop a ML tool for **CROs** that predicts the duration intervals of Phase 3 clinical trials from study protocol data



Minimum Viable Product



MVP Demo



Modeling & Feature Engineering



RandomForest Classification gave the best predictive performance

ClinicalTrials.gov Dataset: completed Phase 3 oncology trials, **n=1,634**

Model	Purpose	Number of Bins	Accuracy	Mean Accuracy K-fold	Precision	MAE
RF	baseline	3	0.245	0.420	0.235	1.137
RF + text features	prediction	3	0.602	0.603	0.592	0.447

We extracted novel features from human-written study protocol fields

Example raw text from study protocol	Output	Feature
<pre>{'measure': 'Difference in mean left anterior descending coronary artery (LAD) mean normal tissue dose (NTDmean) (group B)', 'timeFrame': 'End of radiotherapy (3-4 weeks)'}</pre>	<p>3-4 weeks = 28 days</p>	<p>Time to measure primary AND/OR secondary outcomes</p>
<p>Inclusion Criteria:</p> <ul style="list-style-type: none"> • Complete microscopic excision of early stage invasive ductal or lobular carcinoma (pT1-3b N0-1 M0) of the left breast following breast conservation surgery or mastectomy. • Recommendation for whole breast (groups A and B) or chest wall (Group A only) radiotherapy (with or without tumour bed boost) • Age ≥18 <p>Exclusion Criteria:</p> <ul style="list-style-type: none"> • Requirement for nodal irradiation • Patients with micro- or macro-scopic disease on sentinel node biopsy who have not undergone completion axillary node clearance 	<p>3 criteria</p> <p>2 criteria</p>	<p>Number of inclusion criteria</p> <p>Number of exclusion criteria</p>
<pre>{'measure': 'Overall patient survival rate', 'description': 'The median overall patient survival rate assessed by Kaplan-Meier analysis and log-rank test for treatment comparisons.', 'timeFrame': 'up to 4 years after randomization'}</pre>	<p>True = measuring overall survival</p>	<p>Overall survival outcome</p>

Our novel features were ranked as most important

Feature	Importance	Extracted from Protocol Text?
Time to measure secondary outcomes	0.152	✓
Time to measure primary outcomes	0.138	✓
Number of patients enrolled	0.099	
Number of patient inclusion criteria	0.081	✓
Number of patient exclusion criteria	0.072	✓
Number of study locations	0.071	
Measuring overall survival outcome	0.043	✓

Technical Takeaways



Key learnings

- Study protocol data alone is not sufficient for high accuracy predictions of trial duration
- Using LLMs for feature extraction is difficult with jargon-heavy text
- Our model trained on Phase I data shows promise for outperforming the current best published duration prediction model

Future Work



NLP Feature Extraction

Further explore NLP techniques for feature extraction



User Testing

Get feedback from CRO users



Publication

Publish novel Phase I model findings

Our mission

Improve the quality and efficiency of clinical trials to better deliver novel therapeutic solutions to patients in need

CLINICAL

Acknowledgements

Thank you to Puya & Korin for technical guidance.

Thank you to Stephanie Wong for UX design and branding.

Thank you to our fellow Capstone classmates.

Thank you to our pets for stress relief and emotional support.

Template by Slidesgo.



Nami



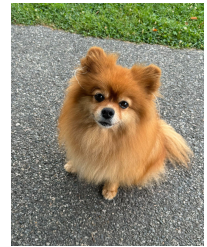
Leo



Frankie



Marco



Lord Lexington

References

1. Armstrong, A. (2023, November 28). Update: Acelyrin points to Cro error that could explain shocking phase 3 failure. Fierce Biotech. <https://www.fiercebiotech.com/biotech/acelyrin-blames-cro-error-shocking-phase-3-failure-sent-shiver-through-biotech-ipo-market>
2. Schmidt, H. (2023, August 11). Nektar sues Eli Lilly for incorrect clinical trial results. PharmaNewsIntelligence. <https://pharmanewsintel.com/news/nektar-sues-eli-lilly-for-incorrect-clinical-trial-results>
3. Long, B., Lai, S. W., Wu, J., & Bellur, S. (2023). Predicting Phase 1 Lymphoma Clinical Trial Durations Using Machine Learning: An In-Depth Analysis and Broad Application Insights. *Clinics and Practice*, 14(1), 69-88.
4. Hutchison, E., Zhang, Y., Nampally, S., Neelufer, I. K., Malkov, V., Weatherall, J., ... & Shameer, K. (2021). Modeling Clinical Trial Attrition Using Machine Intelligence: A driver analytics case study using 1,325 trials representing one million patients. *medRxiv*, 2021-11.
5. Kavalci, E., & Hartshorn, A. (2023). Improving clinical trial design using interpretable machine learning based prediction of early trial termination. *Scientific reports*, 13(1), 121.
6. Wu, K., Wu, E., DAndrea, M., Chitale, N., Lim, M., Dabrowski, M., ... & Zou, J. (2022). Machine learning prediction of clinical trial operational efficiency. *The AAPS Journal*, 24(3), 57.
7. Markey, N., Howitt, B., El-Mansouri, I., Schwartzenberg, C., Kotova, O., & Meier, C. (2024). Clinical trials are becoming more complex: a machine learning analysis of data from over 16,000 trials. *Scientific Reports*, 14(1), 3514.