Evaluating Machine Learning for Forecasting Key Clinical Trial Performance Metrics

Turing - Roche Knowledge Share Series: AI in Clinical Trials

Kevin Wu, PhD Student, Stanford University
Eric Wu, PhD Student, Stanford University
James Zou, Assistant Professor of Biomedical Data Science, Stanford University
Nandini Chitale, Data Scientist, Genentech/Roche
Michael DAndrea, Principal Data Scientist, Genentech/Roche

Agenda

1. Roche/Genentech - Stanford Partnership
2. Measuring Operational Efficiency
3. Data
4. Methods
5. Results
6. Discussion and Conclusion

AI for Health Partnership between Genentech/Roche and Stanford

Goal: develop ML/AI to make clinical trials more effective and to improve precision medicine.

Liu et al. *Nature* 2021

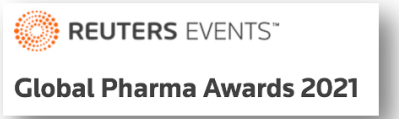
Article | Published: 07 April 2021

Evaluating eligibility criteria of oncology trials using real-world data and AI

Ruishan Liu, Shemra Rizzo, Samuel Whipple, Navdeep Pal, Arturo Lopez Pineda, Michael Lu, Brandon Arnieri, Ying Lu, William Capra, Ryan Copping & James Zou

Nature **592**, 629–633(2021) | Cite this article

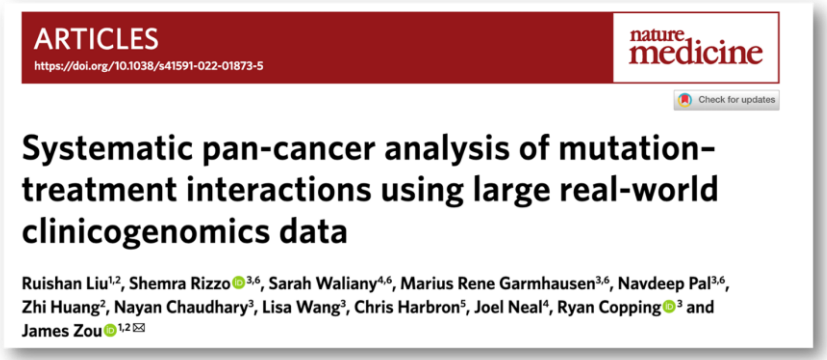
TOP 10
CLINICAL RESEARCH
ACHIEVEMENT AWARDS
PRESENTED BY CLINICAL RESEARCH FORUM



REUTERS EVENTS™

Global Pharma Awards 2021

Liu et al. *Nature Medicine* 2022



ARTICLES

https://doi.org/10.1038/s41591-022-01873-5

nature medicine

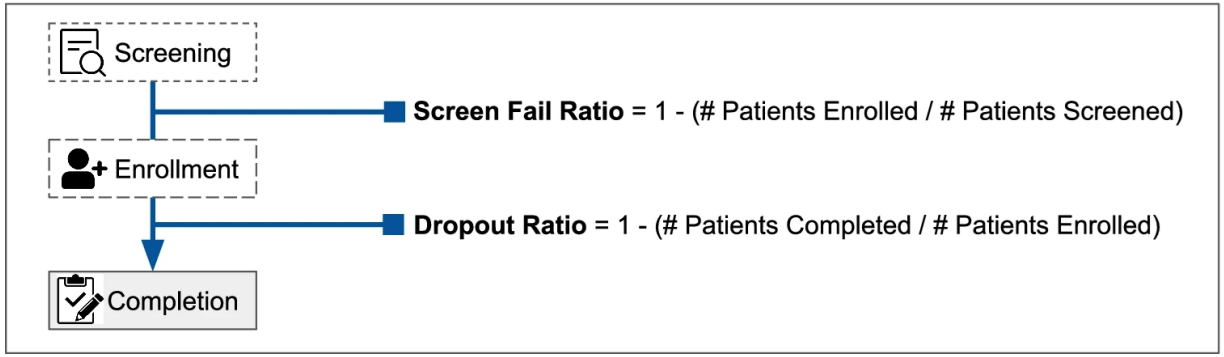
Check for updates

Systematic pan-cancer analysis of mutation–treatment interactions using large real-world clinicogenomics data

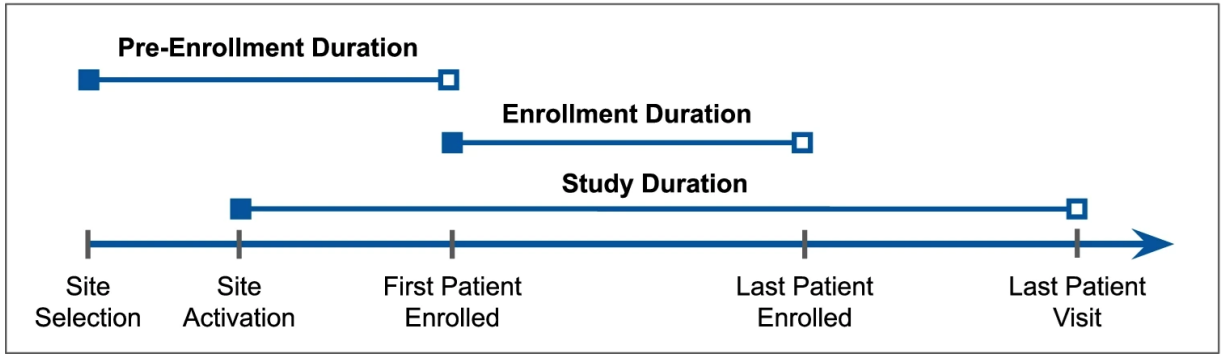
Ruishan Liu[1,2], Shemra Rizzo[3,6], Sarah Waliany[4,6], Marius Rene Garmhausen[3,6], Navdeep Pal[3,6], Zhi Huang[2], Nayan Chaudhary[3], Lisa Wang[3], Chris Harbron[5], Joel Neal[4], Ryan Copping[3] and James Zou[1,2]

Measuring Operational Efficiency

Patient Recruitment



Trial Duration



Data

2,051
Completed trials

2009-2020
Range of starting dates

288
Unique drugs

219
Unique indications

Experimental design
Study phase
Details of planned procedures
23
Operational features
Therapeutic area
# of eligibility eriteria
# of endpoints

On average per trial:

- **11.4** inclusion criteria
- **15.3** exclusion criteria
- **3.9** countries

Methods

Goal

Predict operational efficiency of clinical trials using only **operational features** defined before the trial.

Method: We developed a separate model for each of the **six** operational efficiency metrics:

- ➔ Screen failure ratio
- ➔ Dropout ratio
- ➔ Pre-enrollment duration
- ➔ Enrollment duration
- ➔ Overall trial duration

Methods

Model

We used a **tree-based regression** model (LightGBM) trained with *quantile loss* produce **90% prediction intervals.**

Quantile Loss

$$L_\gamma(y, y^p) = \sum_{i=y_i<y_i^p} (\gamma - 1).\,|y_i - y_i^p| + \sum_{i=y_i \geq y_i^p} (\gamma).\,|y_i - y_i^p|$$

Methods

Evaluation

We use concordance index (**c-index**) to measure the quality of our predictions

Concordance Index

A c-index of **1** indicates the model can correctly predict the order of all of the true labels. Conversely a c-index of **0.5** indicates the model does no better than random.

$$c = \frac{\sum_{i \in U} \left\{ \sum_{T_j > T_i} 1_{f_j > f_i} \right\}}{\sum_{i \in U} \left\{ \sum_{T_j > T_i} 1 \right\}}$$

where U: a set of uncensored data
T_i: an observed survival time of sample i
f_i: a predicted survival time of sample i
$1_{a>b}$: 1 if $a > b$, and 0 otherwise.

Results

We present the ability to predict operational efficiency through **four analysis**:

1. Across **therapeutic area** and **study phase**
2. Across **drug names**
3. Across **years**
4. Prediction **R^2** and Mean Absolute Error (**MAE**)

Results

Prediction across… **therapeutic area** and **study phase**

	Overall	Therapeutic area (C-index)				Study phase (C-index)			
Efficiency metric	**C-index**	**I2O**	**Neuroscience**	**Oncology**	**Other**	**I**	**II**	**III**	**IV**
Screen failure ratio	0.801	0.795	0.765	0.789	0.808	0.622	0.788	0.802	0.771
Dropout ratio	0.791	0.750	0.651	0.715	1.000	0.784	0.801	0.804	0.771
Pre-enrollment duration	0.705	0.724	0.635	0.611	0.687	0.675	0.565	0.587	0.597
Enrollment duration	0.706	0.680	0.709	0.683	0.672	0.764	0.692	0.647	0.609
Trial duration	0.728	0.644	0.766	0.624	0.756	0.808	0.656	0.610	0.666
Average	0.746	0.719	0.705	0.684	0.784	0.731	0.700	0.690	0.683

Results

Prediction across... **drug names**

Validation on unseen Roche drugs (C-index)	Training drug set (*N* = 339)	Testing drug set (*N* = 359)
Screen failure ratio	0.781	0.712
Dropout ratio	0.757	0.738
Pre-enrollment delay	0.674	0.634
Enrollment duration	0.673	0.665
Trial duration	0.699	0.679
Average across metrics	0.717	0.686

Results

Prediction across... **years**

Validation across time (C-index)	Trials completed 2009–2012 (*N* = 439)	Trials completed 2012–2020 (*N* = 376)
Screen failure ratio	0.742	0.726
Dropout ratio	0.630	0.682
Pre-enrollment delay	0.673	0.680
Enrollment duration	0.711	0.669
Study duration	0.704	0.717
Average	0.692	0.695

Results

Additional analyses:

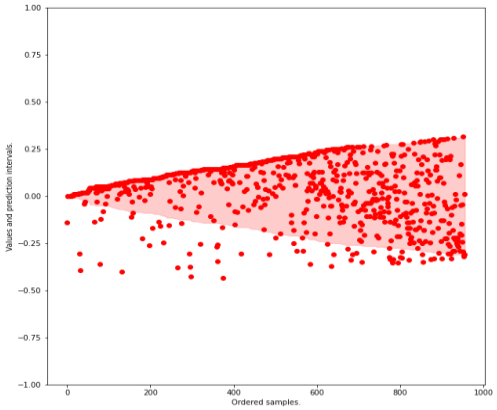
1. Evaluation of **prediction intervals**
2. **Interpretation** of model results

Results

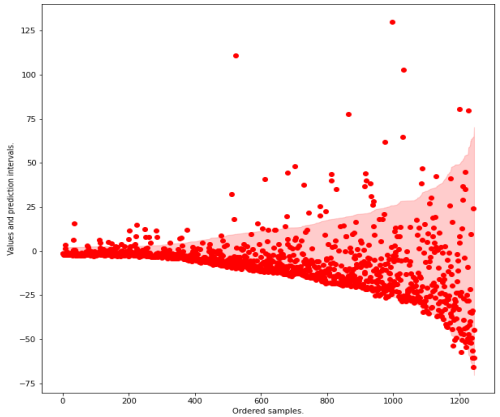
Evaluation of **prediction intervals**

Dot contained within shaded region **->** model prediction interval **contains true value**

Screen failure ratio



Enrollment Rate



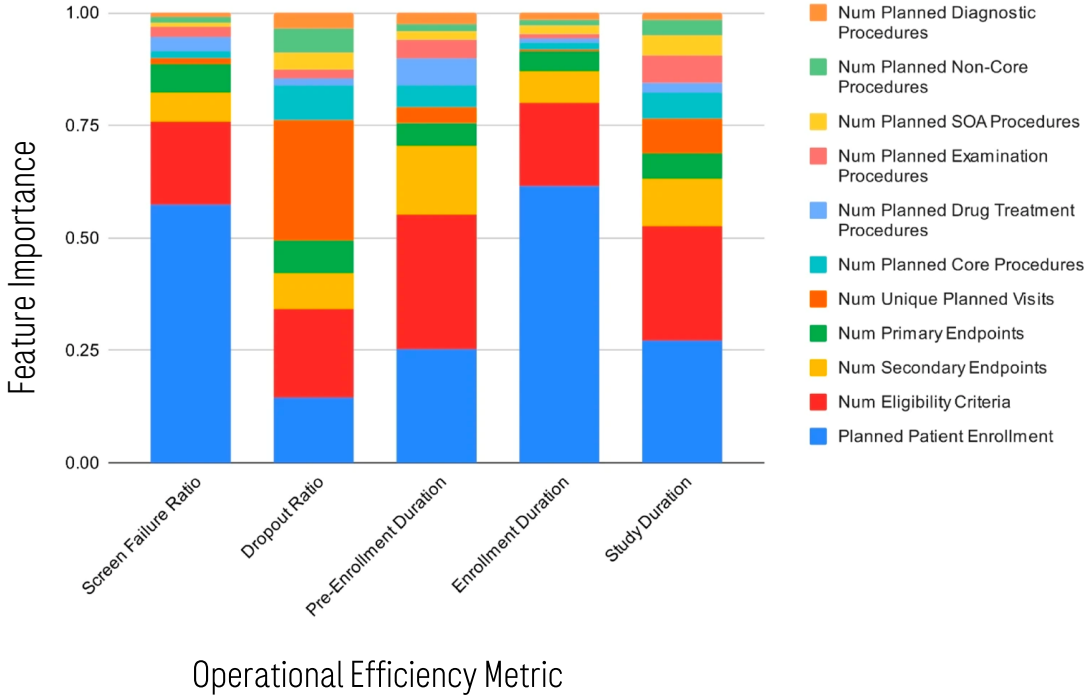
X-axis: Samples in order of prediction interval width, Y-axis: Metric (Mean-centered)

Results

Interpretation of model results

Feature importances of the tree-based model

1. **Patient enrollment** is a large scaling factor for operational efficiency
2. **Eligibility criteria and endpoints** may influence operational efficiency
3. **Procedures** across a variety of categories collectively influence model predictions



Results

Interpretation of model results

How does each **individual operational feature** effect efficiency?

➔ Train a linear regression on the data and report coefficients

Trial Operational Feature	Screen Failure Ratio	Dropout Ratio	Pre-Enrollment Duration	Enrollment Duration	Study Duration
Num Primary Endpoints	0.0064 **	ns	ns	ns	ns
Num Secondary Endpoints	0.0046 ***	ns	ns	-7.2121 **	ns
Number Planned Countries	0.0036 ***	ns	1.2799 **	-7.9442 ***	10.2753 **
Num Eligibility Criteria	ns	ns	ns	1.5514 *	ns
Num Planned Examination Procedures	ns	0.0114 **	ns	ns	ns
Num Planned Non-Core Procedures	-0.0029 *	ns	ns	ns	ns
Num Unique Planned Visits	ns	0.0024 ***	-0.2941 *	0.9518 *	3.7001 ***
Planned Patient Enrollment	ns	ns	ns	0.0164 **	ns

Discussion

- Ability to predict operational complexity robust to therapeutic area, phase, drug, and year
- Screening success and dropout ratio is most easily predicted
- Individual features have significant correlations with operational efficiency outcomes

Acknowledgements

Roche

Michael DAndrea
Nandini Chitale
Melody Lim
Marek Dabrowski
Klaudia Kantor
Hanoor Rangi
Marius Garmhausen
Navdeep Pal
Chris Harbron
Shemra Rizzo
Ryan Copping

Stanford

James Zou
Kevin Wu
Eric Wu
Ruishan Liu