Predicting Clinical Trial Success

Team: Richard Do, Nathan Maton

In partnership with: Yamina Hakem, Brian Lucena, and Jason Moore from

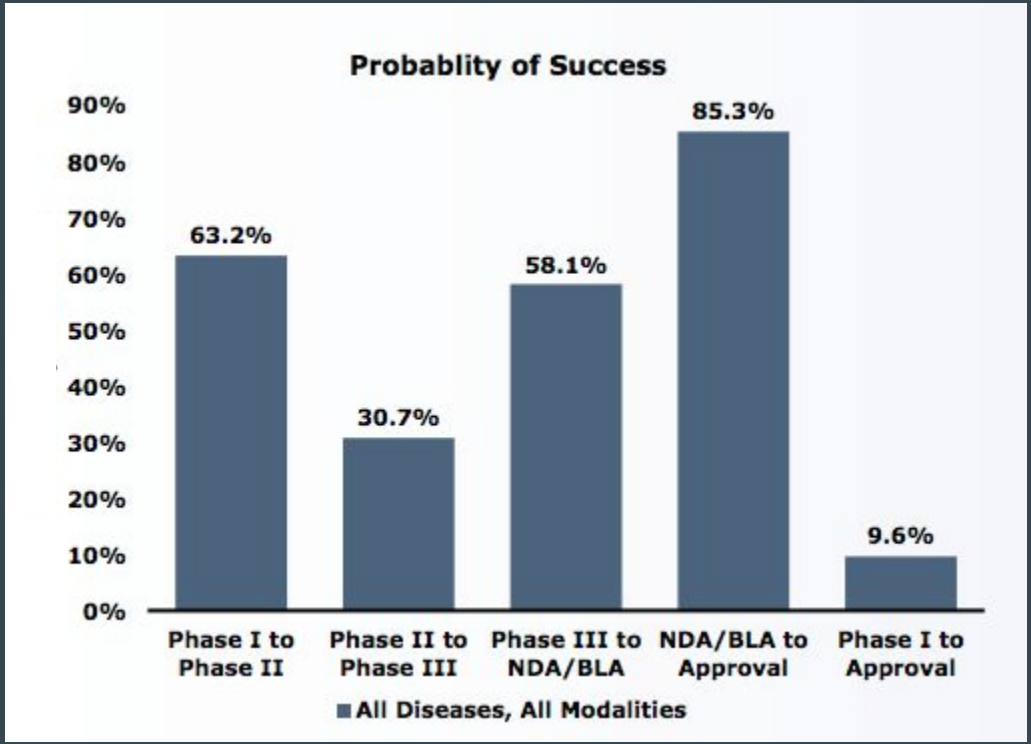
GlobalReach BI

Turning data into profits

How would you make a $2.6 billion decision?

Source

Industry Standard Uses Basic Probability of Success

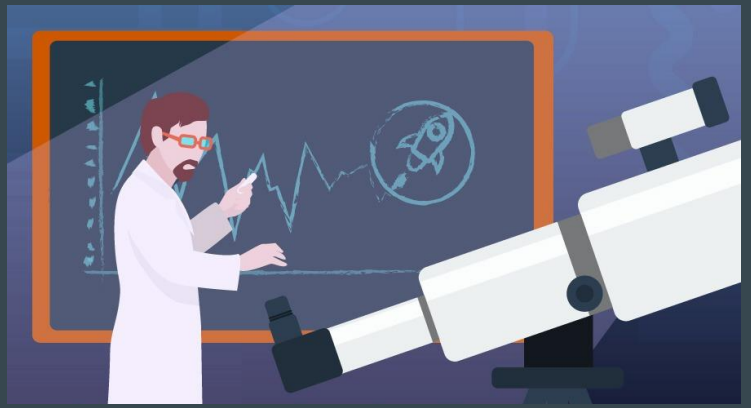


Source: Biomedtracker 2006-2015 report

Problem Statement: Predict Success and Time to Market

Can we predict:

1. Whether a drug will succeed?
2. Which drug will be first to market and quantify the time?



Strategy - Focus on Breast Cancer

Clean trials data

Extract drug features from ADIS Insight trial database

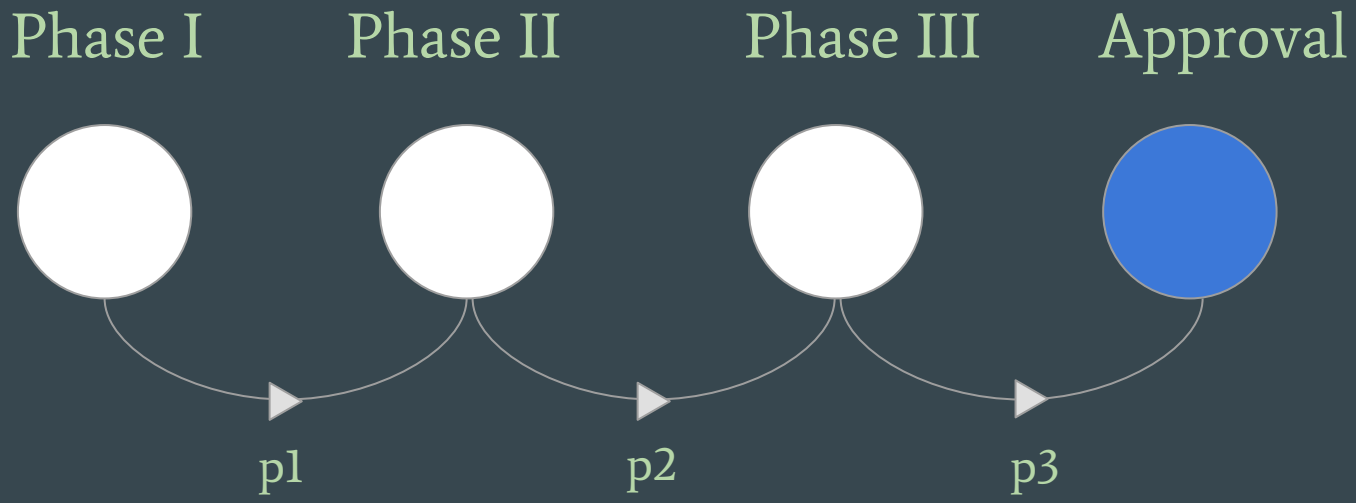
Model success

Use markov chains with each trial phase getting its own state

Test in industry

Work with industry partners and assess value

Simple Markov Chain Explanation



Probability of Success

Engineer Features

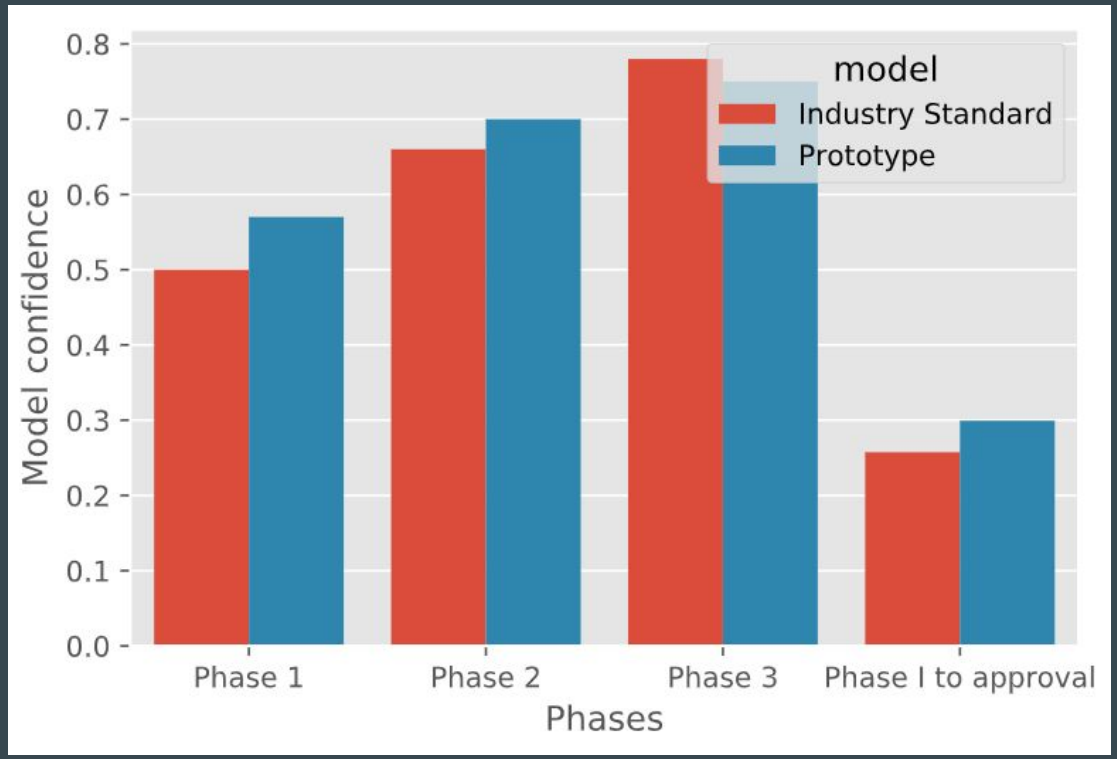
- Subjects in each phase
- Organizations for drug
- Trial centers
- Trial length
- Trial design (12 types, e.g. Randomized)

Question 1: Predict Drug Success

Drivers of Success

	# of Subjects	# of Orgs	Prospective Trial Design
Phase I	Neg	Pos	-
Phase II	Neg	Pos	Pos
Phase III	Pos	Neg	Pos

Model Improvement Over Industry Standard by 14%



Impact:

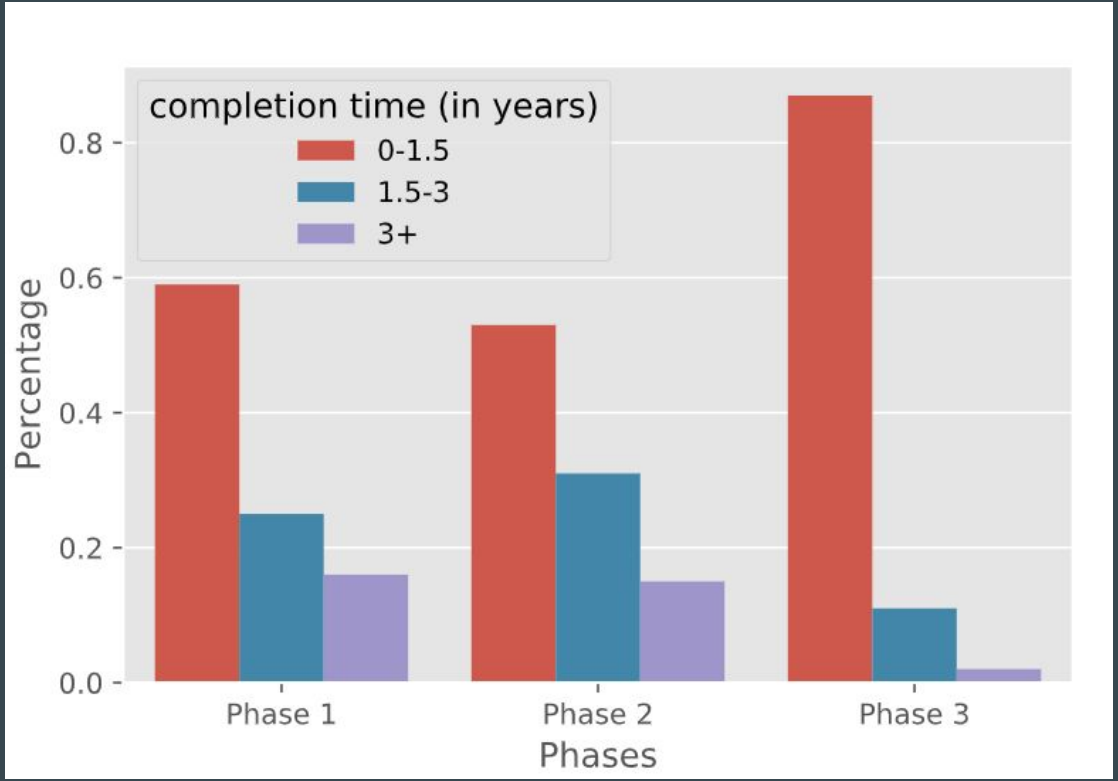
360+ million

on average, cost savings for drug development

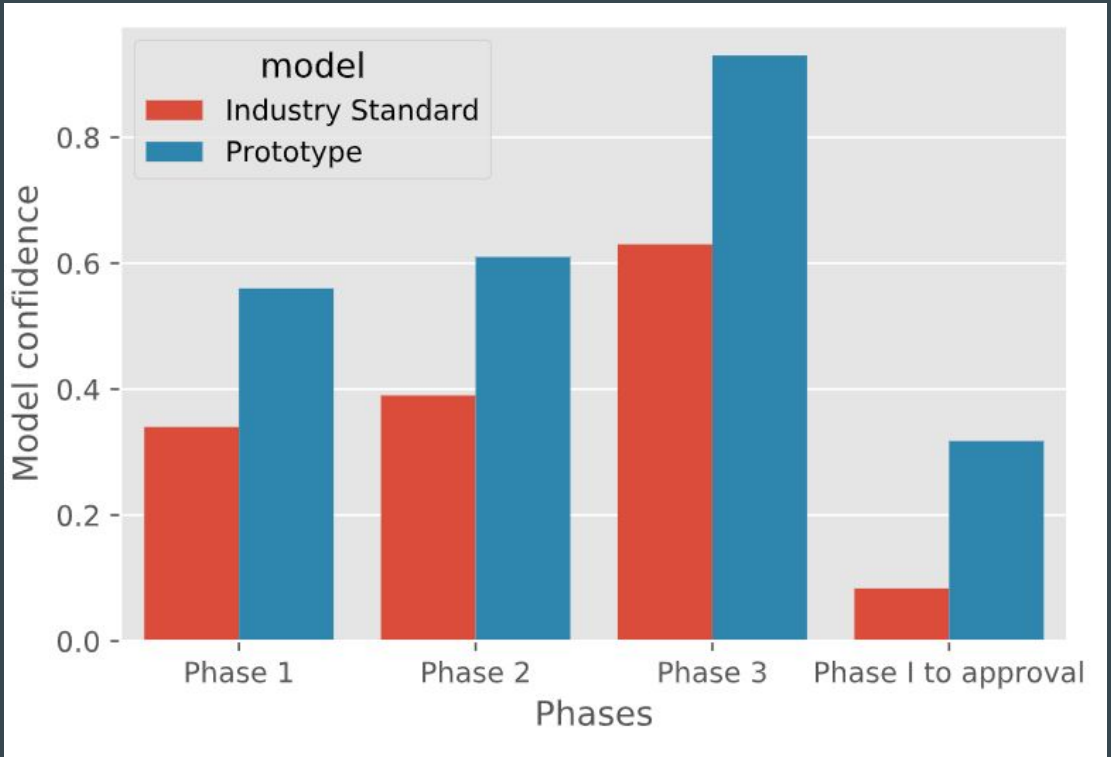
Assumption: Phase success is independent

Question 2: Predict First to Market

Expected Trial Lengths Vary Per Phase



Predicted Phase Completion Better than Industry Standard



Impact:

+12% market share

for first drug to market, on average (300+ million/year)

(source)

Assumption: Phase length is independent

Exciting Results for Partnership Meetings Next Week

- Can make $360+ million, on average, more per drug developed
- First drug to market gains 12% more revenue on average
- Engineering features are the drive to model improvement
- Scale to other diseases

Contact information

Richard Do

/rdooooooo/

in/rdo/

richardddo@gmail.com

Nathan Maton

/NathanMaton/

in/nathanmaton/

nathanmaton@gmail.com

Thanks!

...

Appendix

...

Tools used


python
amazon
web services
Natural Language Analysis
scikit
learn
docker
Seaborn

Key features to understand drivers of trial length

	# of subjects	# of Orgs	Prospective trial design
Phase I			
Phase II			
Phase III		-	

Methodology - Two data sources

Data Available

- **Clinical Trials**: Data about clinical trials to help Breast Cancer starting in 2000
- **Drug data**: Data about drug characteristics in same timeframe