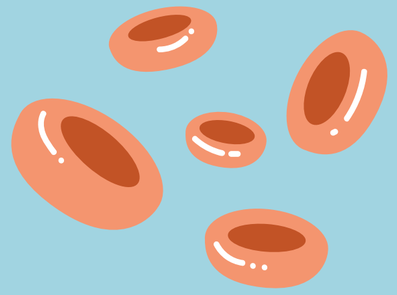
Visualizing and Predicting Trends in Clinical Trials: A Deployed Dashboard Using Clinical Trial Data



Team: Mission Possible
Can Wang, Xiao Wu, Xindi Shan

Background

Exploring and predicting clinical trial trend - why it matters?

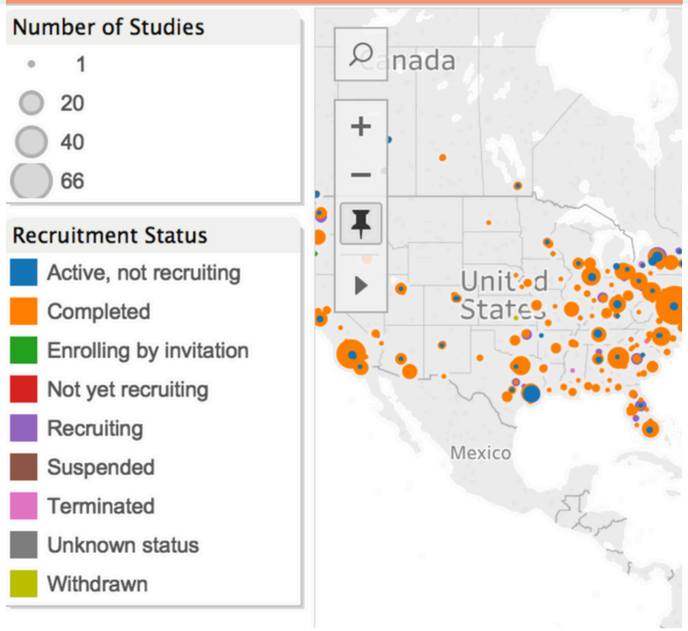
- Identifying patterns aids in **policy-making** and **research prioritization**
- Provide **foresight** for researchers, drug developers and healthcare providers
- Support stakeholders in **regulatory decisions** and **funding allocations**.

ClinicalTrial.gov

- An online database of clinical research studies
- Over 500,000 registered clinical trials across 200+ countries starting from 1982

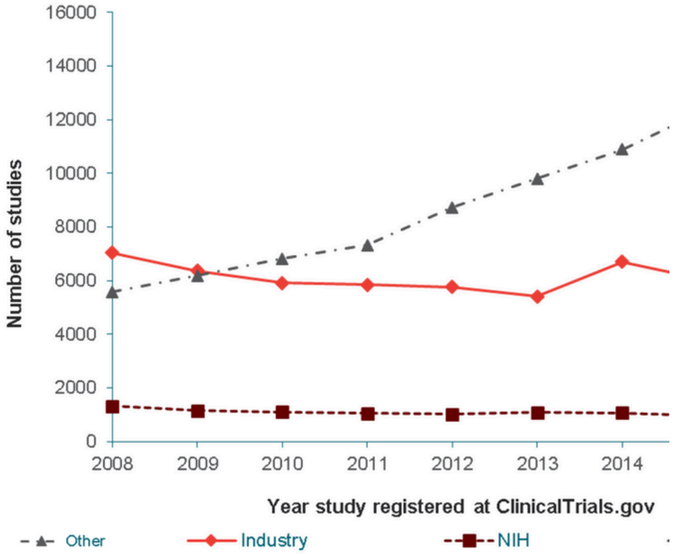
Related Work in Clinical Trial Dashboards

Geographic distribution of EEG-related clinical trials (1)



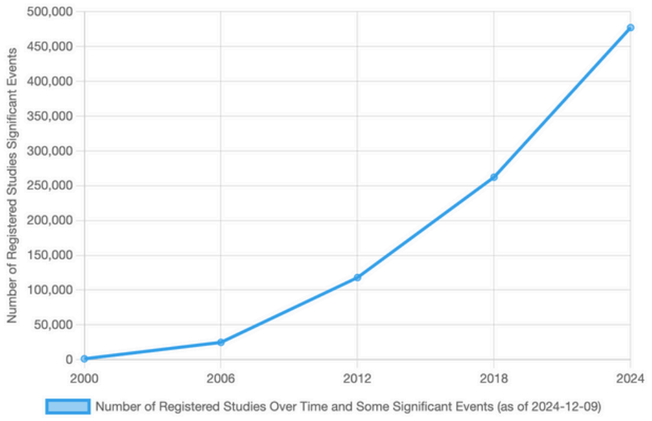
Trends of Interventional Trials 2008-2017 (2)

Study registration summarized by sources of funding



Summary statistics provided by ClinicalTrial.gov (3)

Number of registered studies over time



1. Tibbs, Sheri . 2017. “Where are EEG-related Clinical Trials Being Conducted?.” Ctti-Clinicaltrials.org. 2017. https://aact.ctti-clinicaltrials.org/use_cases/3.
2. Karen Chiswell. 2022. “GitHub - Ctti-Clinicaltrials/Aact: Improving Public Access to Aggregate Content of ClinicalTrials.gov.” GitHub. June 10, 2022. https://github.com/ctti-clinicaltrials/aact.
3. “ClinicalTrials.gov.” 2024. Clinicaltrials.gov. https://clinicaltrials.gov/about-site/trends-charts.

Limitations of Existing Tools and Our Approach

Limitations of Current Tools

- Lack of **comprehensive, and user-friendly** tools for dynamic trend exploration.
- Relying on **outdated** data

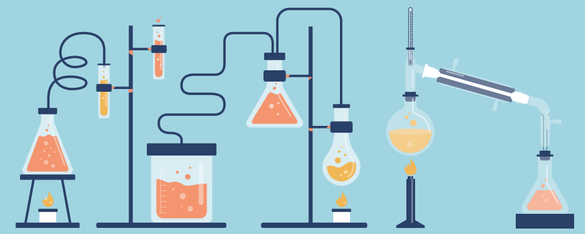
Our Approach

- Using **up-to-date** data from ClinicalTrials.gov
- Developing an **interactive** and **intuitive** dashboard
- Providing **dynamic visualizations** based on:
  - Disease type, trial phase and trial location
- Making **prediction** based machine learning algorithm

Research Goal

What are the trends in the number clinical trials conducted across different countries, disease types and phases, and how do they change over time?

Data Source and Preprocessing



PostgreSQL AACT Database (1)

- Contains all information about every study registered in ClinicalTrials.gov
- Processed and queried using **R**

Data overview

- 51 tables in total
  - Shown on the right is a part of the “Study” table (2)
- Missing values

Study	
nct_id	string
nlm_download_date_description	string
study_first_submitted_date	date
results_first_submitted_date	date
disposition_first_submitted_date	date
last_update_submitted_date	date
study_first_submitted_qc_date	date
study_first_posted_date	date
study_first_posted_date_type	string
results_first_submitted_qc_date	date

Key variables include

- Trial ID, Start Date, Phase, Disease, Countries

Filtering for trials

- Conducted in the **past 10 years**
- **Top 10** most popular **diseases**
- **Top 10** most popular **countries**
- Trials with **no missing value** in selected variables

1. AACT Database | Clinical Trials Transformation Initiative.” n.d. Aact.ctti-Clinicaltrials.org. https://aact.ctti-clinicaltrials.org/
2. Adopted from https://aact.ctti-clinicaltrials.org/schema

Paradigm Integration

- Parallel computing
- Machine learning

Machine Learning Paradigm

data

	nct_id	start_date	phase	year	Breast Cancer	Prostate Cancer	Stroke	Heart Failure	Pain	Obesity	Cancer	Depression	Hypertension	Anxiety	United States	Canada
1	NCT00455039	2023-07-31	PHASE1/PHASE2	2023	1	0	0	0	0	0	0	0	0	0	1	0
2	NCT00573586	2020-12-31	NA	2020	0	1	0	0	0	0	0	0	0	0	0	1
3	NCT00623389	2018-06-01	NA	2018	0	0	1	0	0	0	0	0	0	0	1	0
4	NCT00737893	2017-07-01	PHASE2	2017	0	1	0	0	0	0	0	0	0	0	1	0
5	NCT00908856	2016-01-31	PHASE1	2016	0	0	1	0	0	0	0	0	0	0	1	0
6	NCT00999557	2016-01-31	NA	2016	1	0	0	0	0	0	0	0	0	0	1	0
7	NCT01014000	2015-06-30	NA	2015	0	0	0	1	0	0	0	0	0	0	0	0
8	NCT01123486	2015-01-31	NA	2015	0	0	0	0	1	0	0	0	0	0	1	0
9	NCT01187368	2020-03-31	NA	2020	0	0	0	1	0	0	0	0	0	0	1	0
10	NCT01422096	2018-06-01	EARLY_PHASE1	2018	0	0	0	0	0	1	0	0	0	0	1	0

Data Structure

1. Initial Wide Format
2. Transformation to Long Format
3. Model Matrix Creation

- Time series structure (**year**)
- Simple numeric target (**trial_count**)
- Single category predictor (**country/disease/phase**)

Machine Learning Paradigm

Model: LASSO

WHY LASSO?

- Good for numeric predictions (trial_count)
- Handles time patterns well (trends over years)
- Prevents overfitting through regularization (future predictions)
- Simple to interpret results (stakeholders can understand output)

EFFECTIVITY?

Results Show Success:

- Country model:

R^2 = 0.978; MAE: 22.85 trials

- Disease model:

R^2 = 0.752; MAE: 16.33 trials

- Phase model:

R^2 = 0.986; MAE: 24.50 trials

Code Implementation: **doParallel**

- Ease of use
- Direct compatibility with ***glmnet***
- Works better with ***foreach*** loops (1)

```r
# Setup parallel processing
n_cores <- detectCores() - 1   # Using all cores except one
cl <- makeCluster(n_cores)
registerDoParallel(cl)

# Used in cross-validation
cv_fit <- cv.glmnet(
  x = X_train,
  y = y_train,
  alpha = 1,
  nfolds = 5,
  parallel = TRUE    # Key parameter for parallel processing
)

# Clean up after use
stopCluster(cl)
```

1. University of North Carolina at Chapel Hill. https://unc-libraries-data.github.io/R-Open-Labs/Extras/Parallel/foreach.html

Parallel Computing

EFFECTIVITY

- Reduced **computation time** by distributing work across cores
- Handled **multiple cross-validation** folds simultaneously
- Enabled efficient **lambda** parameter search
- Made training of **three separate models** more efficient

DASHBOARD

- Panel 1: Interactively display trial counts trends in line graph by disease, country and phase.
- Panel 2: Interactively predict trial counts in the future by disease/country/phase.
- Panel 3: Interactively display filtered trial records.

Clinical Trials Dashboard

Select Category

Disease

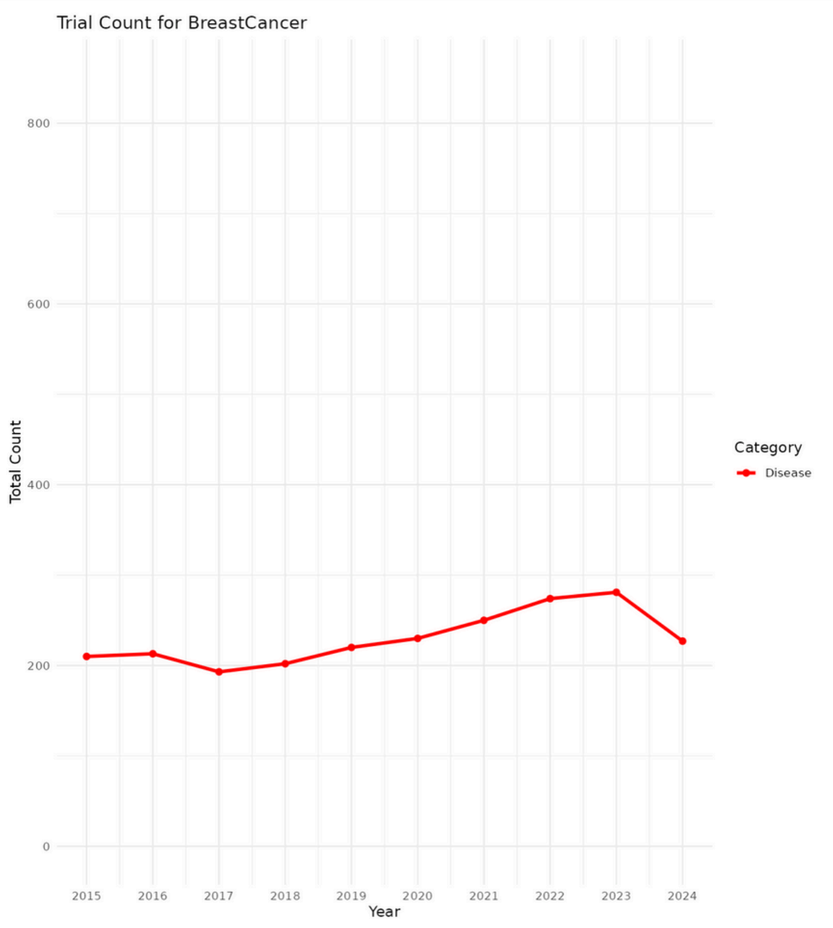
Choose category

Select Disease Label

BreastCancer

Further choose disease/country/phase

Trial Count Over Time



Select Category

country

Select Country

United.States

Select Year for Prediction

2025

Choose year

Predict Trials

click to predict

Prediction Results

```
Predicted Trial Count for United.States in 2025 : 746


 Model Information for country :
 Best lambda: 1.70151896545459
 Test RMSE: 28.7827526449442
 Test MAE: 22.8517325336229
 Test R-squared: 0.978470582319928
```

Information of the model

Filtered Records

Select Year

All

Select Disease Type

All

Choose conditions

Select Country

All

Show 10 entries

Search:

X	nct_id	start_date	phase	year	BreastCancer
1	NCT00455039	2023/7/31	PHASE1/PHASE2	2023	1
2	NCT00573586	2020/12/31	NA	2020	0
3	NCT00623389	2018/6/1	NA	2018	0
4	NCT00737893	2017/7/1	PHASE2	2017	0
5	NCT00908856	2016/1/31	PHASE1	2016	0
6	NCT00999557	2016/1/31	NA	2016	1
7	NCT01014000	2015/6/30	NA	2015	0
8	NCT01123486	2015/1/31	NA	2015	0
9	NCT01187368	2020/3/31	NA	2020	0
10	NCT01422096	2018/6/1	EARLY_PHASE1	2018	0

filtered records

Overview of the dashboard

Clinical Trials Dashboard

Select Category

Disease

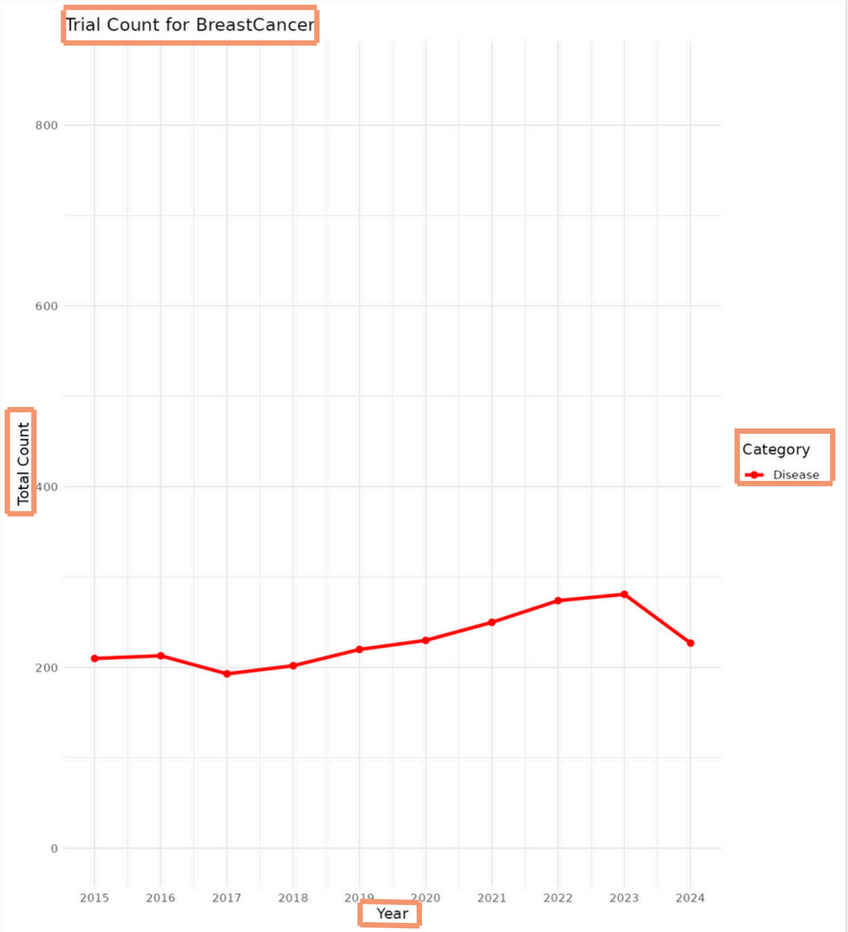
Choose category

Select Disease Label

BreastCancer

Further choose disease/country/phase

Trial Count Over Time



Select Category

country

Select Country

United.States

Select Year for Prediction

2025

Choose year

Predict Trials

click to predict

Prediction Results

```
Predicted Trial Count for United.States in 2025 : 746


 Model Information for country :
 Best lambda: 1.70151896545459
 Test RMSE: 28.7827526449442
 Test MAE: 22.8517325336229
 Test R-squared: 0.978470582319928
```

Information of the model

Filtered Records

Select Year

All

Select Disease Type

All

Choose conditions

Select Country

All

Show 10 entries

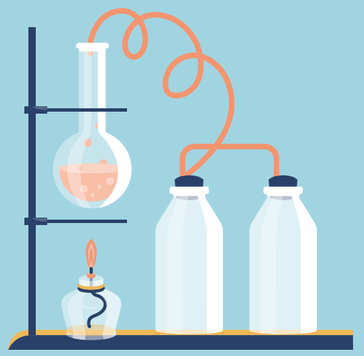
Search:

X	nct_id	start_date	phase	year	BreastCancer
1	NCT00455039	2023/7/31	PHASE1/PHASE2	2023	1
2	NCT00573586	2020/12/31	NA	2020	0
3	NCT00623389	2018/6/1	NA	2018	0
4	NCT00737893	2017/7/1	PHASE2	2017	0
5	NCT00908856	2016/1/31	PHASE1	2016	0
6	NCT00999557	2016/1/31	NA	2016	1
7	NCT01014000	2015/6/30	NA	2015	0
8	NCT01123486	2015/1/31	NA	2015	0
9	NCT01187368	2020/3/31	NA	2020	0
10	NCT01422096	2018/6/1	EARLY_PHASE1	2018	0

filtered records

Overview of the dashboard

Summary of the dashboard



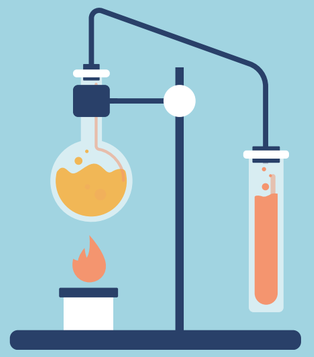
Functionality

Display **descriptive trends** of clinic trials, **predict** future trends and provide **search function** by different conditions.

Usability

The functions are **interactive** and **easy to use**. Figures are clear with titles and legends. Information of predicting models are displayed.

Summary of the dashboard



Originality and complexity

Use up-to-date data, combine the descriptive data and predicting model results, and provide an overview of trends in clinical trials.

Consistency with original goal

We constructed the clinical trial dashboard as planned. However,

- Fewer vairables were used for better visualization and ML model building.
- LASSO model was chosen as the final model.

THANK YOU FOR WATCHING

Q & A?