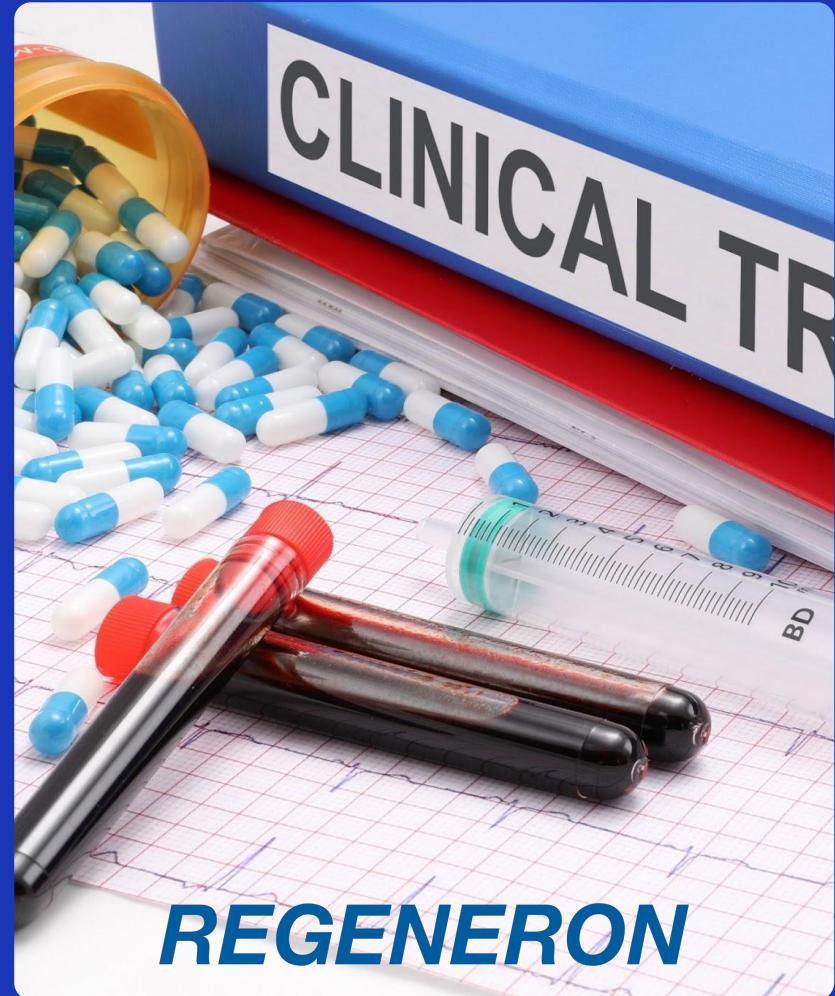


COLUMBIA UNIVERSITY

IEOR E4524

Collected Exhibits from Analytics in Practice teams 2025

Alpha, Bravo and Charlie Teams
collaborating with Regeneron

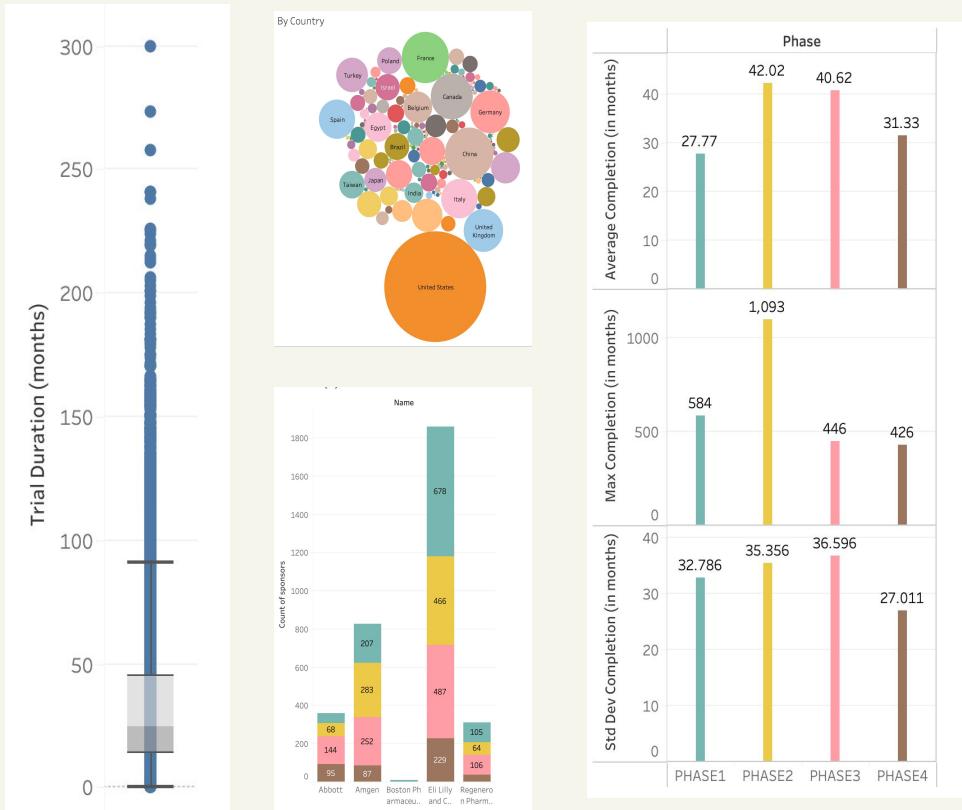


BRAVO TEAM

Visualize ClinicalTrials.gov data to benchmark endpoint complexity

- We created an interactive visualization tool (using data commonly available from ClinicalTrials.gov or derivable from those types of sources) to analyze and benchmark endpoint complexity across clinical trials.
 - The filtering systems allow for a more focused and detailed visualization to further explore connections and insights
- We transitioned from Tableau (free version), to Apache Superset, to Preset.io.

Initial Tableau Visualizations (free version)



Filters and modifications:

- Industry sponsored trials
- Phase 3
- Excluded null values from secondary outcomes, start dates, completion dates
- Converted start and completion dates (from string to date format)
- Created a calculated field to calculate trial duration in months.



Initial Apache Superset Visualizations



- With the free version of Tableau, we struggled with integrating MeSH data with clinical trial data (without an API).
- So we pivoted away from Tableau (free version) to Apache Superset to utilize other features.
 - This solved our API issues.
- However, we still could not collaborate together on this platform and had issues with connecting to the AACT database.
 - Thus, we pivoted away to Preset/back to Tableau (full version?)

Initial Preset Visualizations

The dashboard displays the following data:

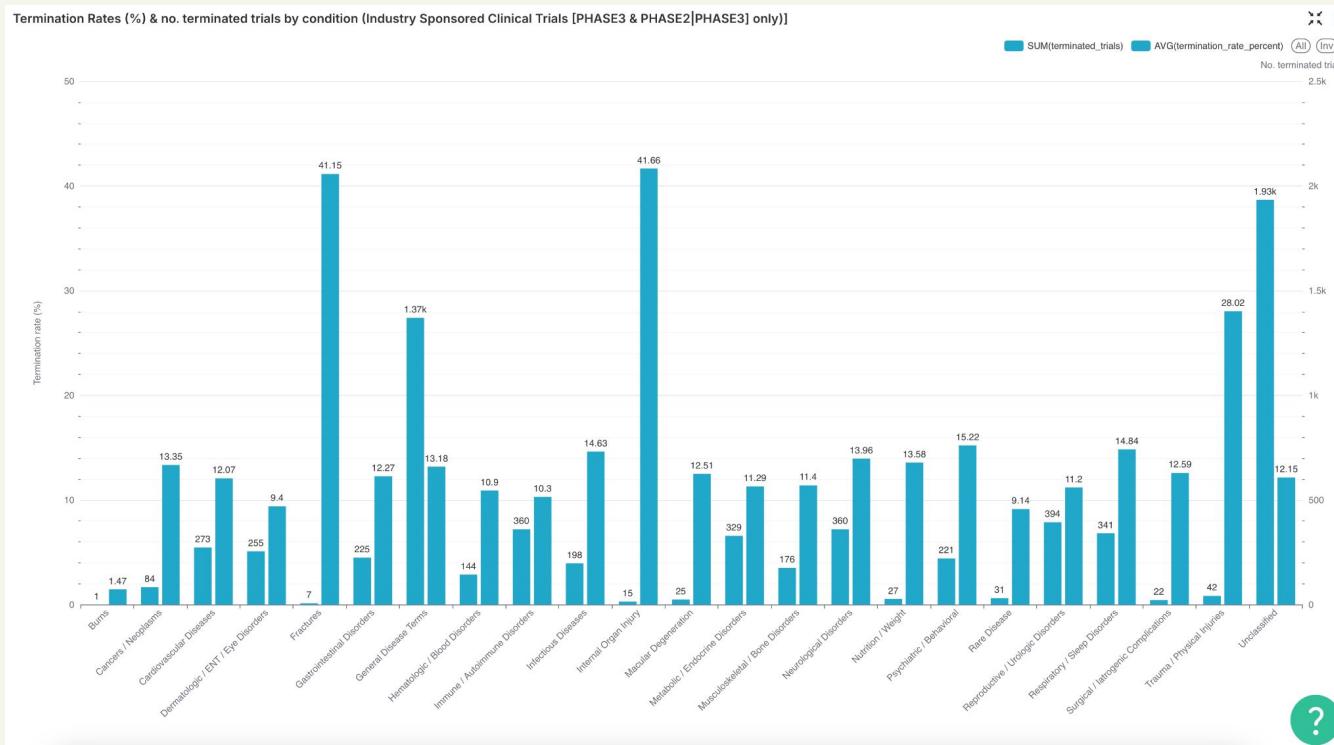
- No. Phase 3 Trials (By Company/Source)**: A table showing the count of Phase 3 trials for various companies and sources.
- Country by Term**: A treemap visualization showing the distribution of terms across different countries.
- Range per Phase**: A scatter plot showing the range of trials per phase.
- by year**: A line chart showing the total number of trials over time.
- Word Cloud Outcome**: A word cloud visualization representing outcomes or concepts.
- intervention based on phases**: A bar chart showing the count of interventions per phase.

Final Preset Visualizations: Dashboard #1 (No Filters)

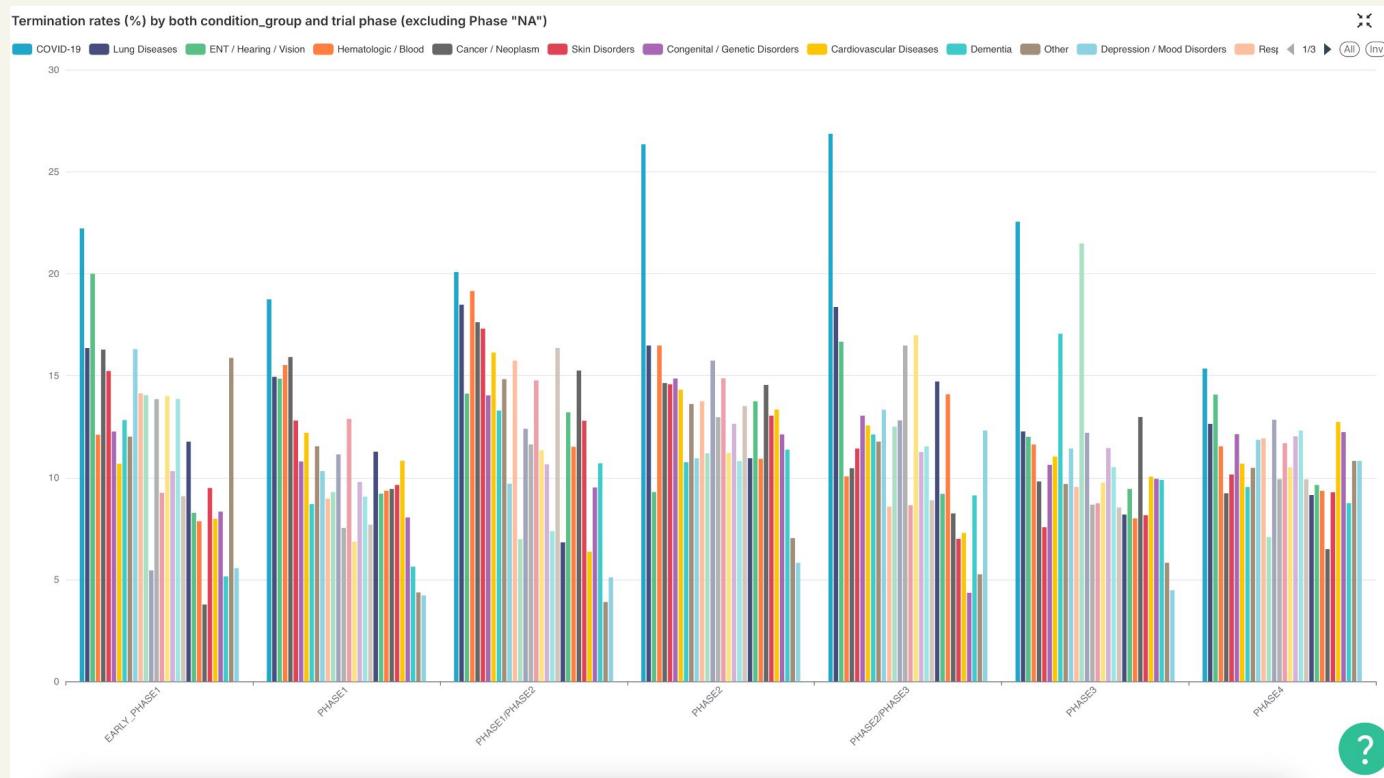
The dashboard displays the following information:

- Termination Rates (%) & no. terminated trials by condition**: A bar chart showing termination rates and counts for various clinical trial conditions.
- Termination rates (%) by both condition_group and trial phase (excluding Phase "NA")**: A grouped bar chart showing termination rates across different trial phases and conditions.
- Active vs Completed Trials Over Time**: A line chart showing the number of active and completed trials over time from 2000 to 2050.
- Number of Trials Missing Results**: A bar chart showing the count of missing results for trials in different phases.
- Study Status Distribution by Country**: A horizontal bar chart showing the distribution of study status (Completed vs Terminated) across countries.
- Primary Outcome Measures**: A section listing various outcome measures such as HbA1C, Sleep quality, Overall Survival Time (OS), and Local Tumor Control.

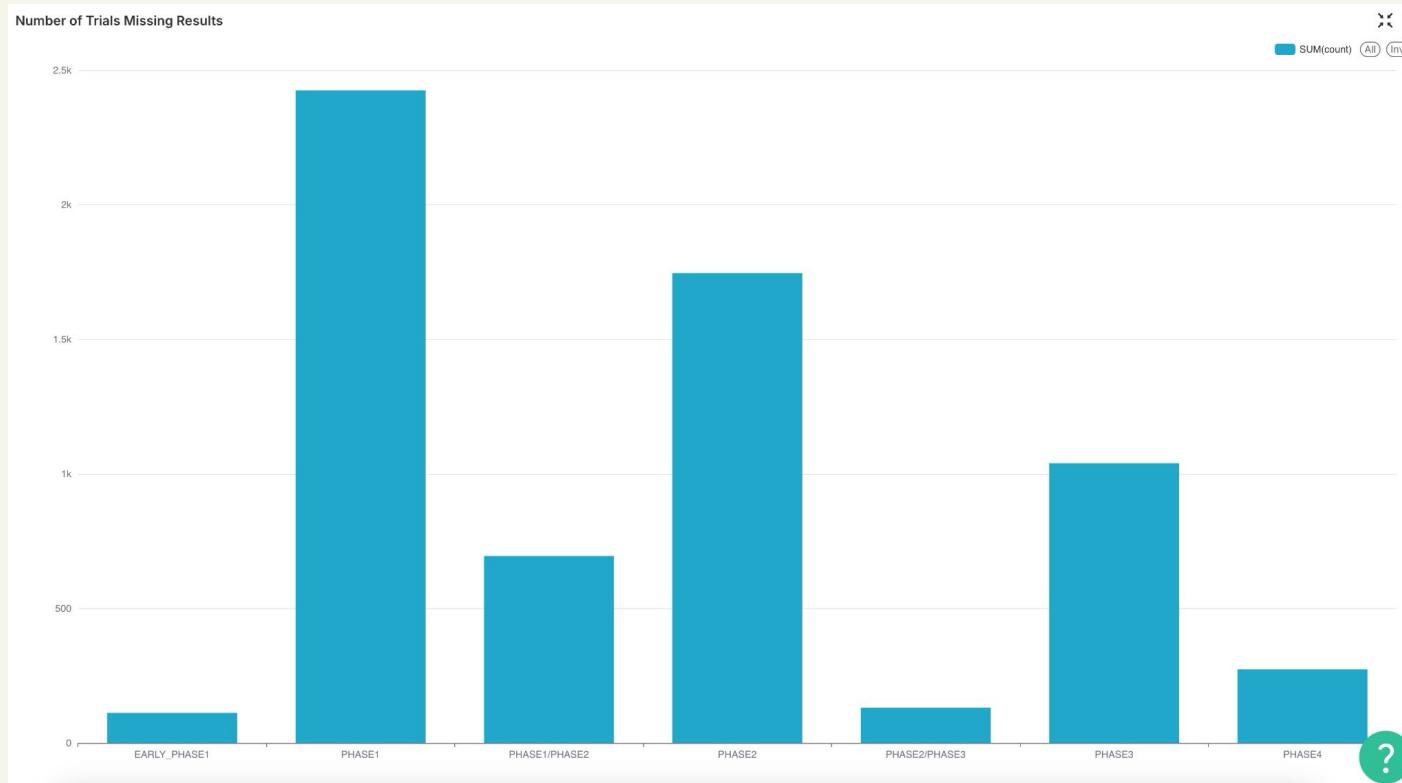
This chart shows termination rates and the number of terminated trials by condition for industry-sponsored Phase 2 and 3 clinical trials, which is important for identifying high-risk therapeutic areas and improving future trial strategies.



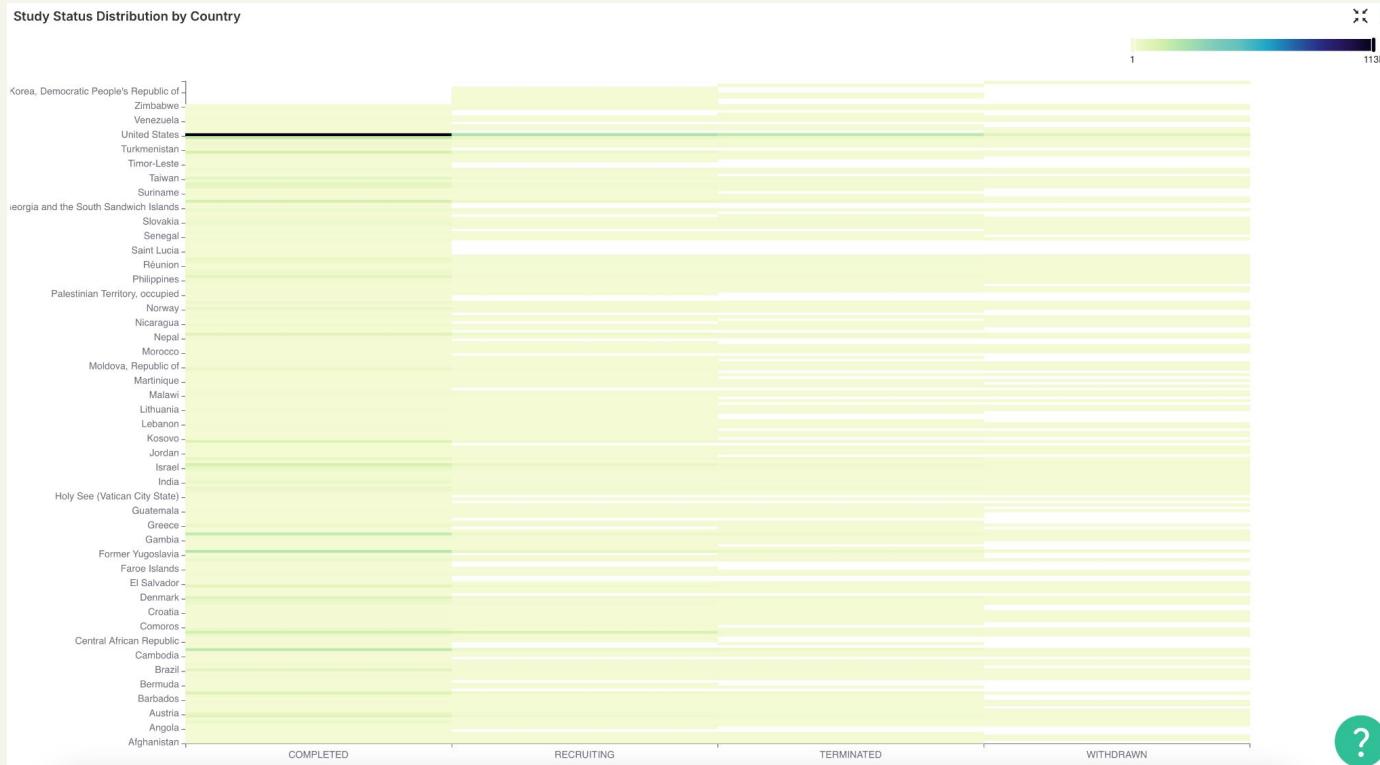
This chart shows termination rates by condition group and clinical trial phase (excluding Phase "NA"), revealing how risks vary across disease areas and stages to guide better resource allocation and trial planning.



This bar chart displays the number of clinical trials missing results by trial phase, highlighting gaps in reported outcomes that can affect transparency and data availability.



This heat map shows the distribution of study statuses across different countries and trial phases, helping to identify geographic and phase-related trends in trial progress.



This word cloud visualizes the most common primary outcome measures across trials, offering a quick overview of frequently studied endpoints.



User Story Dashboard #1

User Story #1

Who: Public Health Researcher

What: Want a dashboard that shows trends in active and completed trials across different conditions.

Why: To evaluate areas of higher trial failure and adjust regulatory planning based on phase-specific risks.



User Story #2

Who: Regulatory Affairs Specialist

What: Want a dashboard that displays termination rates by clinical trial phase and condition.

Why: To ensure compliance and assess the risk factors involved in ongoing trials.

Final Preset Visualizations: Dashboard #2 (With Filters)

The screenshot shows a dashboard interface for 'Clinical Trials Data - Bravo'. At the top left is a 'Filters' button with a gear icon and a dropdown menu labeled 'Conditions' containing '1000 options'. The main title 'Clinical Trials Data - Bravo' includes a yellow star icon, a 'Published' button, a search icon, the author 'Maima Syakhoza', and a timestamp '21 hours'. Below the title, the text 'Clinical Trials Data from clinicaltrials.gov' is displayed. A central section features the 'Bravo Team - Capstone Project' logo with two orange stars and the text 'Columbia Spring 2025'. A detailed description follows: 'This dashboard was developed and is maintained by the Bravo Team, a capstone project group from the Columbia University Master of Science in Business Analytics (MSBA) program during the Spring 2025 semester.' Under 'Team Members:', there is a bulleted list: 'Grace Lee (gl2909@columbia.edu)', 'Maima Syakhoza (mms2374@columbia.edu)', and 'Jess Weng (jw4627@columbia.edu)'. At the bottom, it states: 'This project was supervised by Henry Wei from Regeneron (henry.wei@regeneron.com) and Professor Hardeep Johar from Columbia University (hj2203@columbia.edu)'.

The dashboard for Clinical Trials can be filtered based on conditions (obesity, AIDS, PHA, etc.) It will automatically run the graph based on the filters.

If needed, filters can expanded to other areas such as Phase, Year Started, etc.

Based on the filter, there are currently :
active includes: recruiting, non-recruiting but active, available

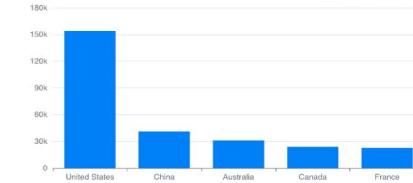
Total Active Clinical Trial

547k

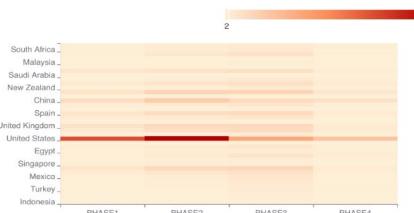
Current Active Clinical Trial

Top 5 Countries with Active Clinical Trials

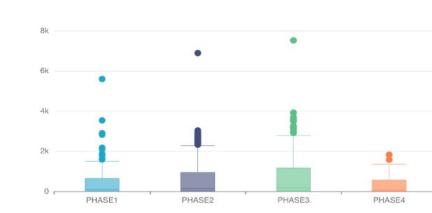
COUNT(*) All Inv



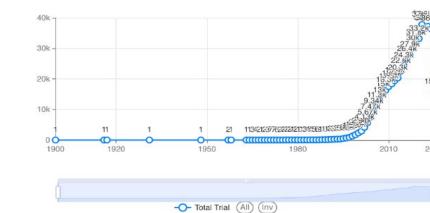
Countries by Phase



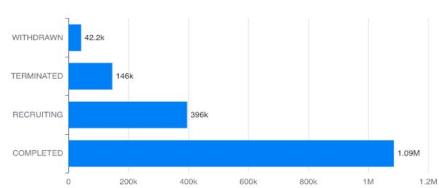
Range per Phase



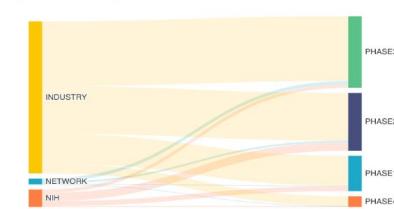
Clinical Trials by Year



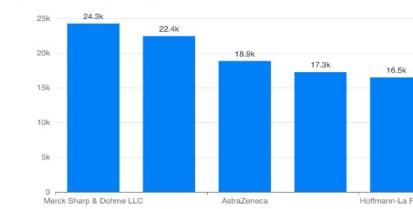
Clinical Trials Status



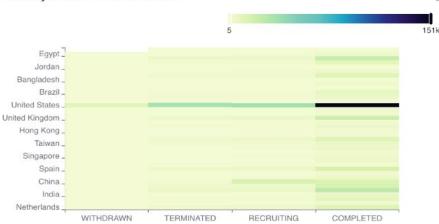
Sponsor Relationship to Phase



Top 5 BioTech Companies Sponsor



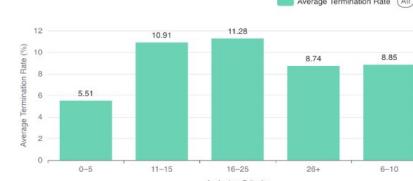
Country with Clinical Trials Status



Empty column

Average Termination Rate Based on Criteria

Average Termination Rate All Inv



The dashboard shows the total current active clinical trials, top 5 countries with the most active clinical trials, countries per phase, time for each phase, research momentum, completed vs terminated trials, sponsors per phase, top 5 sponsors, country per status, and the termination rate per inclusion criteria.

User Story Dashboard #2

User Story #1

Who: Clinical Researcher / Sponsor Organization

What: Want a dashboard that shows clinical trial activity for a specific disease

Why: To strategically plan where and how to launch a successful clinical trial



User Story #2

Who: Biotech Investor

What: Want a dashboard that shows clinical trial landscape

Why: To identify high-growth markets, leading players, and areas with high research momentum to inform investment decisions

Limitations of the Data

1. Unable to get the top countries with a heat map graph. We are able to do it only by filtering manually.
2. Missing financial insights in the clinical trial data.
 - a. Unable to estimate cost per phase or funding amounts due to a lack of cost data
 - b. This would disproportionately impact smaller companies with less resources
3. Missing country-level regulatory or timelines. Unable to get the risk score for each country.
4. No visibility into what protocols succeeded or failed. We are able to get the numbers of inclusion criteria, but can't really compare what makes it terminated.
5. Unable to see the emerging biotech startups, was able to do only the top sponsors.
6. Dashboard can't be shared publicly by link. To share publicly only do it through pdf which makes it not interactive.
7. Termination reasons are vague and often categorized under "other" to conceal other agendas.

Limitations of Preset

Complex Data Relationships

- Complicated joins across the many relationships and tables
 - Ex: AACT is composed of 51 tables that provide information related to clinical trials.

Real-Time Monitoring Challenges

- Not ideal for instant event tracking
 - Preset can auto refresh with like a timer, but it's not event driven (no live feed, no automatic alerts without refreshing).

Statistical Analysis

- Only basic aggregations and summaries
- Advanced calculations must be pre-processed
 - Ex: These calculations must be done through SQL Lab.
- Preset does not support dynamic per-bar sorting of stacked segments (often have to hard code results to make visualizations more intuitive)

Replication of Dashboard

- Specific permissions/passwords are needed to access the database, reducing replicability for others (even if they have access to our code)

Next Steps

Gather User Feedback

- Collect feedback on usability, clarity, and missing insights

Optimize for Performance

- Identify and fix slow-loading charts

Enhancements and Future Features

- Training and Documentations
 - Quick guides: "How to use this dashboard" and "FAQs"
- Scaling the dashboard
 - Plan how to add new diseases, geographies, or trial types