**Goldman Sachs- Clinical Trial Similarity**

***Need:***

Successfully conclude an FDA and National Institute of Health supervised clinical trial proving the efficacy and safety of the proposed treatment

***To do:***

Determine which firms are engaged in similar R&D pipelines

***Existing method:***

Similarity between clinical trials using FDA and NIH Keywords ( based on Medical Subject Headings (MeSH)-controlled vocabulary thesaurus)

***This project:***

develop a new way to form a similarity score between all pairs of clinical studies by only looking at the information contained in the clinical trial and the related application.

***Gains:***

1. Avoid the opportunity for human bias and classification errors
2. Allows similarity to be inclusive of aspects that may be difficult to describe through MeSH keywords.

***Goal:***

By creating a score of similarities between each individual clinical trial and knowing the firm which submitted the trial, we will seek to form an overall similarity between the research priorities of the various submitting pharmaceutical firms.

***Dataset:***

<https://aact.ctti-clinicaltrials.org/>

Study limited to : Investigational New Drug Application (IND)

Helpful Hints: [https://aact.ctti-clinicaltrials.org/points\_to\_consider](https://aact.ctti-clinicaltrials.org/points_to_consider%20)

Database Schema: <https://aact.ctti-clinicaltrials.org/schema>

Data Dictionary: <https://aact.ctti-clinicaltrials.org/data_dictionary>

***Reference:***

1. <https://ieeexplore.ieee.org/document/6867604/>

2. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5977658/>