

ClinicalTrials.gov

STATS 418: Final Project

Isaac Schwarz

Exploratory Data Analysis

Data Source:

- **ClinicalTrials.gov** API
- Initially retrieved ~10,000 study records
- Final dataset included **~100 high-quality records** after cleaning
 - Due to time and technical constraints, a larger sample was not processed

General Observations:

- Most columns were **fairly clean**
- Some columns contained **missing or ambiguous values** (e.g., “Unknown”)
- A subset of **irrelevant variables** was dropped from the analysis

Variables

Feature Variables (10)

- **Categorical (9):**
 - study_type, allocation, intervention_model, masking,
 - primary_purpose, gender, healthy_volunteers,
 - lead_sponsor_class, intervention_types
- **Numerical (1):**
 - minimum_age (parsed from string format)

Target Variables (3):

- death_rate
- serious_adverse_event_rate
- other_adverse_event_rate

Methodology

Model Choice: XGBoost

- Selected for its ability to handle **missing data**, **nonlinear relationships**
- Robust performance with **mixed-type input data** (categorical + numerical)
 - Can handle **categorical variables** natively **without encoding**
- Suitable for **multi-output regression** tasks

Web App

About

This app uses an XGBoost machine learning model to predict the likelihood of adverse events occurring in clinical trials based on trial characteristics. The model was built off of publicly available data from [ClinicalTrials.gov](#)

Predictions include:

- Rate of Deaths
- Rate of Serious Adverse Events
- Rate of Other Adverse Events

Instructions

- Fill in all trial parameters
- Click 'Generate Predictions'
- Review the predicted outcomes

Clinical Trial Adverse Event Prediction

This application predicts adverse events for clinical trials based on study characteristics. Enter the trial parameters below to get predictions for potential adverse events.

Trial Parameters

Study Type

INTERVENTIONAL

Allocation Method

RANDOMIZED

Intervention Model

PARALLEL

Masking/Blinding

NONE

Primary Purpose

TREATMENT

Minimum Age (years)

18

Gender

ALL

Healthy Volunteers

☒ True
☐ False

Lead Sponsor Class

INDUSTRY

Intervention Types

DRUG

Generate Predictions

Prediction Results

Deaths Rate

0.227

Serious Adverse Events Rate

0.042

Other Adverse Events Rate

0.420

Predictions are estimates based on historical data and should be used for planning purposes only.

Clinical Trial Adverse Event Prediction