Abhishek Mehra

ALY 6980 Capstone

Week 4 Assignment 1

Abhishek Mehra

Strategic Insights into Epilepsy Research Funding

mehra.ab@northeastern.edu

**Introduction**

**Strategic Insights into Epilepsy Research Funding**

The main purpose of my individual project is to perform a funding analysis focused on epilepsy for the years 2015-2024. This involves examining trends in funding allocations, sources of funding, and the impact of these financial resources on epilepsy research and treatment outcomes. My analysis aims to identify funding gaps, effectiveness of invested resources, and potential areas for future investment.

The research "Trends and hotspots in gene research of epilepsy in children" by Tian, Y., Zhang, focuses on the increasing number of publications on epilepsy genes in children, with the United States and China leading this trend. The top journals in this field have published highly cited articles, offering valuable insights for researchers. However, the conclusions in these articles should be interpreted with caution, as they reflect the authors' perspectives. Key findings include prominent keywords such as "epilepsy," "Dravet syndrome," "neurodevelopmental disorders," "gene testing," and "whole-exome sequencing." Important knowledge clusters identified were "epileptic encephalopathy," "Dravet syndrome," "focal cortical dysplasia," "rolandic epilepsy," "copy number variation," "ketogenic diet," "monogenic epilepsy," and "PTEN mutation." Recent research hotspots (2019-2022) highlighted "developmental and epileptic encephalopathy (DEE)," "neurodevelopmental disorders," "gene testing," and "whole-exome sequencing." Some topics, such as "epileptic encephalopathy" and "Dravet syndrome," have remained consistently significant throughout the study period.

The individual analysis of epilepsy funding will serve as a critical component of the broader group project, which may involve a comprehensive review of various aspects of epilepsy management and research over the same period. Your detailed funding analysis will provide insights into:

1. **Resource Allocation**: How funds are being allocated across different research and treatment areas in epilepsy.
2. **Effectiveness of Funding**: How effectively the funding has been utilized in achieving tangible improvements in epilepsy care and research.
3. **Strategic Recommendations**: Based on your analysis, you can offer strategic recommendations for future funding strategies, helping to guide more effective resource utilization.

**The dataset: Source:** https://reporter.nih.gov/

The dataset contains detailed information on epilepsy-related projects funded by NIH. Here's a brief overview of some key columns in your dataset:

* **NIH Spending Categorization**: Category of spending by NIH.
* **Project Terms**: Keywords or terms associated with the project.
* **Project Title**: Title of the research project.
* **Public Health Relevance**: Description of the project's relevance to public health.
* **Administering IC**: NIH Institute or Center administering the grant.
* **Award Notice Date**: Date the award was announced.
* **Fiscal Year**: Fiscal year of the funding.
* **Total Cost**: Total cost associated with the project.
* **Funding IC(s)**: Institutes or Centers funding the project.
* **Direct Cost IC** and **InDirect Cost IC**: Direct and indirect costs attributed to the Institutes or Centers.

The data set filtered from. 2015 to 2024 has the following summary statistics.

* **Count**: 10,602 projects
* **Mean**: The average funding per project is approximately $486,234.97.
* **Standard Deviation**: The standard deviation, which indicates the variability of funding amounts, is $777,925.44.
* **Minimum**: The smallest funding amount is $0.00. This might represent projects with no direct NIH funding recorded.
* **25th Percentile (Q1)**: 25% of the projects received $197,002.00 or less in funding.
* **Median (50th Percentile)**: The median funding per project is $363,438.50.
* **75th Percentile (Q3)**: 75% of the projects received $532,737.00 or less in funding.
* **Maximum**: The highest funding for a single project is $24,767,583.00.

**Research Methodology:**

The research methodology involves a quantitative analysis of funding data related to epilepsy research projects. This includes:

* **Data Collection**: Gathering data on epilepsy funding from relevant sources, primarily focusing on NIH funded projects.
* **Data Cleaning and Preparation**: Processing the data to ensure accuracy and usability, which involves handling missing values, filtering out irrelevant data, and transforming data into a suitable format for analysis.
* **Descriptive Analysis**: Calculating statistical measures to understand the distribution, central tendency, and variability of funding amounts. This helps in identifying trends, averages, and outliers in the dataset.
* **Trend Analysis**: Analyzing changes in funding over time to understand growth patterns and potential impacts of external factors on funding allocations.
* **Visualization**: Creating visual representations of data to better communicate findings and insights, facilitating easier interpretation of the data's story.

**Software Platform and Commands:**

The analysis will primarily use Python, a versatile programming language well-suited for data analysis and visualization, along with specific libraries tailored for these tasks:

* **Python**: The core programming language used for scripting and automating the analysis process.
* **Pandas**: A Python library for data manipulation and analysis. Key commands include:
  + **read\_excel()** for importing data from Excel files.
  + **groupby()** and aggregation functions like **sum()**, **mean()**, and **count()** for summarizing data.
  + **describe()** for generating descriptive statistics of datasets.
* **Matplotlib and Seaborn**: These Python libraries are used for data visualization, providing a wide range of plotting functions. Key commands include:
  + **plt.figure()**, **plt.hist()**, **plt.plot()**, **plt.title()**, **plt.xlabel()**, and **plt.ylabel()** from Matplotlib for creating histograms and line charts.
  + **sns.boxplot()** from Seaborn for creating box plots that illustrate distributions with quartiles and detect outliers.

**Results**

| Statistic | Value |
| --- | --- |
| Count | 10,602 projects |
| Mean | $486,234.97 |
| Standard Deviation | $777,925.44 |
| Minimum | $0.00 |
| 25th Percentile | $197,002.00 |
| Median (50th Percentile) | $363,438.50 |
| 75th Percentile | $532,737.00 |
| Maximum | $24,767,583.00 |
|  |  |

Table 1: Summary statistics

Descriptive statistics summarizes the funding amounts and analyze trends over the period from 2015 to 2024. This can include:

* **Total annual funding**: Sum of all project costs per year.
* **Average project cost**: Average cost of projects per year.
* **Number of projects funded per year**: Count of projects awarded each year.

Let's calculate these statistics to provide an initial overview of epilepsy funding trends in your dataset. We will focus on the 'Fiscal Year' and 'Total Cost' columns for this purpose. ​​

Here are the descriptive statistics for epilepsy funding from 2015 to 2023:

| Fiscal Year | Total Funding | Average Project Cost | Number of Projects |
| --- | --- | --- | --- |
| 2015 | $370,677,517.00 | $360,581.24 | 1028 |
| 2016 | $455,702,109.00 | $443,289.99 | 1028 |
| 2017 | $500,459,743.00 | $470,799.38 | 1063 |
| 2018 | $558,543,977.00 | $485,690.41 | 1150 |
| 2019 | $580,702,458.00 | $496,326.89 | 1170 |
| 2020 | $589,718,547.00 | $512,798.74 | 1150 |
| 2021 | $642,823,078.00 | $551,306.24 | 1166 |
| 2022 | $635,978,126.00 | $529,540.49 | 1201 |
| 2023 | $615,231,899.00 | $539,677.10 | 1140 |

Table 2 : Summary Statistics Table: Funding of Epilepsy research

| Administering Institute | Total Funding | Average Funding per Project | Total Direct Cost | Total Indirect Cost |
| --- | --- | --- | --- | --- |
| AHRQ | 1,445,617 | 180,702.10 | 980,248 | 465,369 |
| CIT | 369,698 | 184,849.00 | 0 | 0 |
| CLC | 0 | 0.00 | 0 | 0 |
| FDA | 3,193,256 | 1,064,419.00 | 2,587,186 | 606,070 |
| FIC | 12,592,683 | 307,138.60 | 5,240,966 | 617,094 |
| NCATS | 77,468,576 | 1,075,952.00 | 17,268,070 | 6,348,865 |
| NCCDPHP | 5,334,143 | 762,020.40 | 3,869,913 | 989,332 |
| NCCIH | 36,305,643 | 648,315.10 | 39,263,960 | 13,550,241 |
| NCI | 28,289,962 | 304,193.10 | 15,035,870 | 6,636,167 |
| NEI | 80,330,412 | 347,750.70 | 46,793,850 | 24,264,408 |
| NIMH | 822,761,116 | 539,161.90 | 502,851,500 | 212,894,311 |
| NIMHD | 1,591,049 | 318,209.80 | 974,842 | 515,628 |
| NINDS | 2,746,603,859 | 408,781.60 | 1,870,600,000 | 733,074,240 |
| NINR | 18,739,761 | 506,480.00 | 10,903,150 | 3,882,275 |
| NLM | 5,043,364 | 315,210.20 | 3,216,096 | 1,227,268 |

Table 3: Summary Statistics Table: Funding by Administering Institute

**Observations:**

* There is a general trend of increasing total funding from 2015 to 2021, peaking in 2021, with a slight decrease in the following years.
* The average project cost also shows an upward trend, with some fluctuations, indicating an increase in either the scale or the cost of funded projects over time.
* The number of projects funded each year remains relatively stable, except for a noticeable drop in 2024. This could be due to incomplete data for that year or changes in funding strategies.

A graph of a graph with a line and a red line

Description automatically generated with medium confidenceFig : Total funds and funding in years 2014-2023

| Funding Mechanism | Number of Projects |
| --- | --- |
| Non-SBIR/STTR | 8,337 |
| Other Research-Related | 1,197 |
| Training, Individual | 911 |
| Research Centers | 607 |
| Intramural Research | 336 |
| SBIR/STTR | 322 |
| Training, Institutional | 171 |
| Other | 46 |
| R and D Contracts | 29 |
| Interagency Agreements | 6 |
| Construction Grants | 2 |

Fig :Distribution of Projects by Funding Mechanism

The distribution of epilepsy-related projects by funding mechanism is summarized in this table. It shows a large majority of projects are funded through Non-SBIR/STTR mechanisms, emphasizing the traditional research funding routes in contrast to smaller categories such as SBIR/STTR and research centers.

A graph with blue squares

Description automatically generated

Fig : Top 10 Administering ICs by Total Funding

The horizontal bar chart displays the total funding allocated by the top ten NIH Institutes and Centers (ICs) for epilepsy research. NINDS leads with a substantial margin, reflecting its primary role in neurological research funding. This visualization aids in understanding how resources are distributed among different NIH components and highlights where the majority of research support is concentrated.

A graph with blue squares

Description automatically generated

Fig: Project Counts by Funding Mechanism

This bar chart represents the distribution of epilepsy-related projects across various NIH funding mechanisms. It clearly shows that the majority of projects are supported through Non-SBIR/STTR funding mechanisms, followed by other research-related grants and individual training grants. The chart serves as a useful tool for visualizing the diversity and focus of funding strategies employed by NIH to support epilepsy research.

In addition to the year-wise summary of funding data, I have performed a trend analysis on the most common words in the research. This analysis provides insights into the evolving priorities in epilepsy research by examining the focus areas and themes in epilepsy research projects from 2014 to 2023 using natural language processing techniques.

A close-up of words

Description automatically generatedFig : Word Cloud of most common words in studies

**Observations**

**Project Titles:**

* **Epilepsy**: Remains a dominant theme, highlighting the focus on this condition. This consistent emphasis indicates ongoing efforts to address the complexities and challenges associated with epilepsy.
* **Neuro, brain, seizures**: Indicative of research centered around neurological aspects and seizure activity. These terms reflect the focus on understanding the brain's role in epilepsy and developing strategies to manage and prevent seizures.

**Project Terms:**

* **Seizures, epilepsy, neuro**: Reflects a strong emphasis on understanding and treating seizures and epilepsy. This suggests that much of the research is aimed at finding effective treatments and understanding the neurological basis of epilepsy.
* **Mechanisms, brain, treatment**: Suggests significant research on the mechanisms of epilepsy and its treatment methods. Researchers are not only looking at the clinical aspects but also delving into the biological and physiological processes that underlie epilepsy.

A graph of different colored lines

Description automatically generated with medium confidence

Yearly **Trends in Project Terms**: This chart tracks the frequency of key terms such as 'neurological', 'research', 'clinical', 'epilepsy', and 'study' within the project terms over the years. It highlights fluctuations and possibly emerging focuses in the research domain.

Yearly **Trends in Project Titles**: Similarly, this chart examines terms like 'epilepsy', 'neurological', 'treatment', 'seizure', and 'therapy' in project titles over the years, offering insights into the evolving research themes and areas of emphasis.

**Forecast**

A graph showing a line of a graph

Description automatically generated with medium confidence

Forecast for Total Cost

The forecasted total cost for the year 2024 based on historical funding trends is approximately **$655.46 million**.

This forecast is made using a simple linear regression model based on funding data from the previous years. It provides a straightforward prediction of future funding, assuming similar conditions and trends continue.

**Forecasted Trends for the Term "Epilepsy"**

Here are the forecasted frequencies for the term "epilepsy" over the next three years:

**In Project Terms**

* **2024**: Approximately 1617 mentions
* **2025**: Approximately 1609 mentions
* **2026**: Approximately 1602 mentions

**In Project Titles**

* **2024**: Approximately 208 mentions
* **2025**: Approximately 210 mentions
* **2026**: Approximately 211 mentions

These forecasts suggest a relatively stable trend for the term "epilepsy" in both project terms and titles over the next few years, with a slight decline in the project terms and a slight increase in project titles. This stability might indicate sustained interest and ongoing research focus on epilepsy.

**Forecasting Rule/Equation/Algorithm:**

1. **Linear Regression**:
   * Used for forecasting the total funding based on historical funding data.
   * **Equation**: y=mx+by = mx + by=mx+b
     + Where yyy is the forecasted funding, mmm is the slope of the line (rate of change in funding), xxx is the fiscal year, and bbb is the intercept.
   * This simple model assumes a linear relationship between the year and the total funding.
2. **ARIMA Model for Time Series Forecasting**:
   * Used for forecasting the term frequency of "epilepsy".
   * **ARIMA(p, d, q)**: Where p is the order of the autoregressive part, d is the degree of first differencing involved, and q is the order of the moving average part.
   * The ARIMA model was configured as ARIMA(1, 1, 1) for both project terms and titles, indicating a basic level of autoregression and moving average with a single differencing to address non-stationarity in the time series data.

### Recommendation for Sponsor

To effectively raise funding for epilepsy research, Grik Therapeutics should focus on leveraging individual training and research-related funding mechanisms, particularly through grants like R01, R21, and R35/R37. Emphasize partnerships with leading NIH Institutes such as NINDS, NIMH, and NICHD, and utilize strong keywords like "brain disease" and "surgical/non-surgical" to enhance visibility. Collaborate with top principal investigators, including Lori L. Isom, William A. Catterall, Huajun Feng, and Christopher Werley, to strengthen research credibility. Additionally, increase connections and network extensively, ensuring the research is backed by peer reviews and endorsements from top PIs and research institutes to solidify funding opportunities.

**References**

Privitera, M. (2006, January 1). *Large clinical trials in epilepsy: Funding by the NIH versus pharmaceutical industry*. Epilepsy Research. [https://doi.org/10.1016/j.eplepsyres.2005.09.020Links to an external site.](https://doi.org/10.1016/j.eplepsyres.2005.09.020)

Alpern, J. D., Zhang, L., Stauffer, W. M., & Kesselheim, A. S. (2017). Trends in Pricing and Generic Competition Within the Oral Antibiotic Drug Market in the United States. *Clinical Infectious Diseases*, *65*(11), 1848–1852. <https://doi.org/10.1093/cid/cix634>

Ghosh, S., Sinha, J. K., Ghosh, S., Sharma, H., Bhaskar, R., & Narayanan, K. B. (2023). A Comprehensive Review of Emerging Trends and Innovative Therapies in Epilepsy Management. Brain Sciences, 13(9), 1305. https://doi.org/10.3390/brainsci13091305